



# STIC Search Report

## Biotech-Chem Library

### STIC Database Tracking Number

TO: Georgia L Helmer  
Location: 9d14 / 9e12  
Friday, November 07, 2003  
Art Unit: 1638  
Phone: 308-7023  
Serial Number: 09 / 622500

From: Jan Delaval  
Location: Biotech-Chem Library  
CM1-1E07  
Phone: 308-4498

jan.delaval@uspto.gov

### Search Notes

Edward Hart submitted your jobs to queue on November 7, 2003.

I have printed the results for you. Any questions should be directed to Edward at your convenience next week.

*pending  
removed*

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:50:29 ; Search time 1314 Seconds

(without alignments)  
4883.091 Million cell updates/sec

Title: US-09-622-500b-3

Sequence: 264  
1 atgatacaagatgagagatcaaa.....gccctggtcccccgaagtaa cca

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 1258238056 residues

Word size : 3

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:  
1: em\_estba:  
2: em\_esthum:  
3: em\_estim:  
4: em\_estnu:  
5: em\_estrov:  
6: em\_estropl:  
7: em\_estro:  
8: em\_hic:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hic:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estfun:  
16: em\_estom:  
17: em\_gss\_hum:  
18: em\_gss\_inv:  
19: em\_gss\_pln:  
20: em\_gss\_vrt:  
21: em\_gss\_fun:  
22: em\_gss\_mam:  
23: em\_gss\_mus:  
24: em\_gss\_pro:  
25: em\_gss\_rtd:  
26: em\_gss\_phg:  
27: em\_gss\_vrl:  
28: gb\_gss1:  
29: gb\_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	8.0	474	13	BQ280354
2	19	7.2	136	28	A2601419
3	19	7.2	497	14	C8220913
4	19	7.2	656	10	BG705752

5	19	7.2	681	10	B5370873
6	19	7.2	727	28	AQ544745
7	19	7.2	766	29	B2571110
8	19	7.2	889	13	B0857584
9	19	7.2	942	12	B1078258
10	19	7.2	970	29	CNSC0731
11	19	7.2	1111	29	B2577558
12	19	7.2	1122	29	A4489437
13	19	7.2	140	9	A726470
14	19	6.8	140	9	A7940990
15	19	6.8	140	9	A7940990
16	19	6.8	140	9	A7940990
17	19	6.8	140	9	A7940990
18	19	6.8	140	9	A7940990
19	19	6.8	140	9	A7940990
20	19	6.8	140	9	A7940990
21	19	6.8	140	9	A7940990
22	19	6.8	140	9	A7940990
23	19	6.8	140	9	A7940990
24	19	6.8	140	9	A7940990
25	19	6.8	140	9	A7940990
26	19	6.8	140	9	A7940990
27	19	6.8	140	9	A7940990
28	19	6.8	140	9	A7940990
29	19	6.8	140	9	A7940990
30	19	6.8	140	9	A7940990
31	19	6.8	140	9	A7940990
32	19	6.8	140	9	A7940990
33	19	6.8	140	9	A7940990
34	19	6.8	140	9	A7940990
35	19	6.8	140	9	A7940990
36	19	6.8	140	9	A7940990
37	19	6.8	140	9	A7940990
38	19	6.8	140	9	A7940990
39	19	6.8	140	9	A7940990
40	19	6.8	140	9	A7940990
41	19	6.8	140	9	A7940990
42	19	6.8	140	9	A7940990
43	19	6.8	140	9	A7940990
44	19	6.8	140	9	A7940990
45	19	6.8	140	9	A7940990

## ALIGNMENTS

RESULT 1  
BQ280354  
LOCUS  
DEFINITION  
BQ280354.1 GI:20523060  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Mickelson, J.R.  
Veterinary Pathobiology  
University of Minnesota  
1988 Fitch Avenue, University of Minnesota, St. Paul, MN 55106, USA  
Tel: 612 624 1246  
Fax: 612 625 0204  
Email: mickel@tc.umn.edu  
Seq primer: M13 Reverse  
Location/Qualifiers  
1. 474

```

/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/sex="Male"
/clone_lib="Canine Brain cDNA Library"
/notes="Organ: Brain; Vector: pSPORT1 (Gibco BRL); Site_1:
NotI; Site_2: SalI; Tissue was taken from the frontal,
occipital, temporal and parietal lobes, olfactory bulb,
hippocampus, cerebellum, thalamus, hypothalamus, midbrain
, pons, and medulla."
BASE COUNT      121 a      115 c      167 g      60 t      11 others
ORIGIN

Query Match      8.0%; Score 21; DB 13; Length 474;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

146 TGGACGAGGGCCAGCCCGCCT 166
|||||
61 TGGACGAGGGCCAGCCCGCCT 61

RESULT 2
AZ601419      306 bp      DNA      linear      GSS 13-DEC-2000
LOCUS
DEFINITION
M0419L09R Mouse 10kb plasmid UGCCIM library Mus musculus genomic
clone UGCCIM0419L09 R, genomic survey sequence.
ACCESSION
AZ601419
VERSION
AZ601419.1 GI:11723609
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 306)
Dunn,D., Royagui,A., Barber,X., Seacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished
Contact: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5636
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0419 row: L column: 09
Seq primer: CACACGAGAAACAGCATATGACC
Class: plasmid ends
High quality sequence scop: 306.
Location/Qualifiers
1..306
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCCIM0419L09"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/notes="Vector: pMD22rv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/charters/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were

```

```

ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi:473211419b/AF199072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      84 a      66 c      96 g      60 t
ORIGIN

Query Match      7.2%; Score 19; DB 28; Length 306;
Best Local Similarity 100.0%; Pred. No. 2,2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 CAGGCCAGCCAGCCCGAG 36
|||||
138 CAGGCCAGCCAGCCCGAG 156

RESULT 3
CB220913      497 bp      mRNA      linear      EST 10-FEB-2003
LOCUS
DEFINITION
1AB027H07 Bos taurus Abomasum #1 library Bos taurus cDNA, mRNA
sequence.
ACCESSION
CB220913
VERSION
CB220913.1 GI:26291427
KEYWORDS
EST.
SOURCE
Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 497)
Hansen,C., Fu,A., Meng,Y., Li,C., Okunev,E., Sensen,C.W., Gordon
,P.V.K. and Moore,S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished
Contact: Dr. Stephen Moore
Reef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/Fac, Dept of AFNS, U of A, Edmonton, AB, T6G 2P6, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ua.bertha.ca
Insert Length: 497 Std Error: 0.00
POLY(A)=NC.
Location/Qualifiers
1..497
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Smooth muscle"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XL1-B. JEMRI"
/clone_lib="Bos taurus Abomasum #1 library"
/notes="Organ: Abomasum; Vector: Uni-Z24PKS; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT      136 a      124 c      174 g      63 t
ORIGIN

Query Match      7.2%; Score 19; DB 14; Length 497;
Best Local Similarity 100.0%; Pred. No. 2,5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

146 TGGACGAGGGCCAGCCCGC 64
|||||
75 TGGACGAGGGCCAGCCCGC 93

```

```

RESULT 4      BC705752      656 bp      mRNA      linear      EST 07-MAY-2001
LOCUS         BC705752
DEFINITION    Homo sapiens CDNA clone IMAGE:4791574 5'.
ACCESSION     BC705752
VERSION       BC705752.1 GI:13980409
KEYWORDS
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS       NIH-MGC http://mgc.nhl.nih.gov/.
TITLE         Unpublished
JOURNAL
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-remail@nih.gov
              Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
              cDNA Library Preparation: Michael C. Brownstein (NHGR-1), Shiroki
              Toshiyuki and Piero Carninci (RIKEN)
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA sequencing by: Invitae Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/URL at:
              http://image.llnl.gov
              Plate: LHAM:0668 row: f column: 23
              High quality sequence stop: 542.
              Location/Qualifiers
                1..656
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4791574"
                  /tissue_type="hypothalamus"
                  /lab_host="CH10B"
                  /clone_1ib="NIH_MGC_96"
                  /note="Organ: Brain; Vector: pBluescript (modified
                    pluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI; gcgcag
                    1'; Oligo-dT primed using primer 5'-TTTCTTTTTTTT-TVN-3',
                    size-selected for average insert size 2.3 kb and
                    normalized to RCF 5. This is a primary library enriched
                    for full-length clones and constructed using the
                    Cap-trapper method [Carninci, in preparation]. Library
                    constructed by M. Brownstein (NIH/NHGR National
                    Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT    144 a      200 c      256 g      106 t
ORIGIN
Query Match   7.2%; Score 19; DB 10; Length 656;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY            6 CAAGCGGAGATCAAGCCC 24
               |||:|||||:|||||:
Db            469 CAAGGTGGAGATCAAGCC 487
RESULT 5
LOCUS         BE370873
DEFINITION    Mus musculus cDNA clone IMAGE:358046 5'.
ACCESSION     BE370873
VERSION       BE370873.1 GI:9316236
KEYWORDS
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              (bases 1 to 681)
              NIH-MGC http://mgc.nhl.nih.gov/.

```

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-rcmail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.Y.A.G.E. Consortium (LINU)
            DNA sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.Y.A.G.E. Consortium/LINU at:
            http://image.lim.nih.gov
            Plate: L1468752 row: c column: 23
            High quality sequence start: 54
            High quality sequence stop: 544.
            Location/Qualifiers
                source
                    1..681
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6"
                        /db_xref="taxon:10090"
                        /clone="IMAGC:3588046"
                        /tissue_type="spontaneous tumor, metastatic to mammary."
                        /stem_cell_origin="
                            /cell_line="FHL2B"
                            /cell_line_id="NCI-GAP_0029"
                            /note="Organ: lung; Vector: pCMV-Script6; Site: 1 (SalI);
                            Site 2: NotI; Cloned unidirectionally. Primer: 01-30 dt.
                            Library constructed by Life Technologies. Investigator
                            providing samples: Gilbert Smith, NIH"
BASE COORDINATE      149 a      163 c      235 g      134 t
ORIGIN
Query Match      7.2% Score 19; DB 10; Length 681;
Best local Similarity 100.0%; Pred. No. 2,7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
CY      134 TGAAGATCACCTGGACGA 152
          TTTTTTTTTTTTTTTT
DB      61 TGAAGATCACCTGGACGA 79
RESULT 6
A0544745      727 bp      DNA      linear      GSS 28-MAY-1996
LOCUS LOCUS
DEFINITION C17B1-E-2611G7.FF C17B1-E: Homo sapiens genomic clone 2611G7,
genomic survey sequence.
ACCESSION      A0544745
VERSION      A0544745.1 GI:4903556
KEYWORDS      GSS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 727)
Zhao,S., Adams,M.D., Niernan,W., Malek,J., Satyana,H., Simon,M. and
Venner,J.C
Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
Map Building
Unpublished
Other GSSes: C17B1-E-2611G7.FP
Contact: Shanying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20852
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@wglr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M3-21
Class: BAC ends.
Location/Qualifiers

```

```

source
1.727
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="261.G7"
/sex="male"
/cell_type="sperm"
/clone_lib="GIBI-E1"
/notes="Vector: pGEOBAC1; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT
210 a 146 c 186 g 185 t

Query Match
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7.2%; Score 19; DB 29; Length 727;

QY 20 AGCCGAGCCAGCCGCGT 38
|||||
Db 557 AGCCGAGCCAGCCGCGT 575

RESULT 7
B2571110 766 bp DNA linear GSS 17-DEC-2002
JOCUS
DEFINITION
msb2_1733.x1 msh Pseudomonas aeruginosa genomic clone msh2_1733,
genomic survey sequence.
B2571110
B2571110.1 GI:27236171
GSS.
SOURCE
Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 766)
REFERENCE
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burris,O.L., Kaul,R. and Olsen,X.V.
Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
JOURNAL
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: ckraymond@u.washington.edu
COMMENT
Class: shotgun.
FEATURES
Location/Qualifiers
1..766
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msb2_1733"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT
123 a 176 c 184 g 283 t

Query Match
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7.2%; Score 19; DB 29; Length 766;

QY 126 CGTGTGTGTAAGATCACC 144
|||||
Db 161 CGTGTGTGTAAGATCACC 179

RESULT 8
BUB57584/c
LOCUS BUB57584 889 bp mRNA linear EST 16-OCT-2002

```

```

DEFINITION
AGENCOURT 104735.7 NIH-MGC 107 Homo sapiens cDNA clone
IMAGE:5647348 5', mRNA sequence.
ACCESSION
BU857584
VERSION
BU857584.1 GI:24042576
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 889)
REFERENCE
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (JUN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILN at:
http://image.nih.gov
Date: DEC2987 row: 1 column: 24
High quality sequence scop: 596.
FEATURES
Location/Qualifiers
1..889
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6647448"
/tissue="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH-MGC 107"
/notes="Organ: Breast; Vector: pOT7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene, and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library."
BASE COUNT
141 a 321 c 285 g 140 t 2 others

Query Match
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7.2%; Score 19; DB 13; Length 889;

QY 18 CAGGCCGACGACGCCCCAG 36
|||||
Db 789 CAGGCCGACGACGCCCCAG 771

RESULT 9
BI078258 942 bp mRNA linear EST 20-JUN-2001
LOCUS BI078258
DEFINITION
60287259CF1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5034263 5',
mRNA sequence.
ACCESSION
BI078258
VERSION
BI078258.1 GI:14496565
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 942)
REFERENCE
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.

```

CDNA Library Arrayed by: The J.M.A.G.E. Consortium (LINT)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: WeG clone distribution information can be  
 found through the J.M.A.G.E. Consortium/LINT at:  
<http://image.lim.gov>  
 Plate: LLM11042 row: d column: 24  
 High quality sequence stop: 787.  
 Location/Qualifiers

## FEATURES

source

1..942  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5034263"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 morchs"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Salt;  
 Site 2: Not; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"  
 232 a 230 c 314 g 166 t

## BASE COUNT

232 a 230 c 314 g 166 t

## ORIGIN

Query Match 7.2% Score 19; DB 12; Length 942;  
 Best Local Similarity 100.0%; Pred. No. 2,9e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

132 GGTGAAGATCACCCTGGAC 150

## Db

10 GGTGAAGATCACCCTGGAC 28

RESULT 10  
 CENS0713L/c 970 bp DNA linear GSS 06-JUL-2001  
 LOCUS 77 end of clone AZ0AA3-4C12 of library AZ0AA from strain CBS 712 of  
 DEFINITION Kluyveromyces marxianus, genomic survey sequence.  
 ACCESSION AL424423  
 VERSION AL424423.1 GI:12207617  
 KEYWORDS GSS.  
 SOURCE Kluyveromyces marxianus  
 ORGANISM Kluyveromyces marxianus  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 1 (bases 1 to 970)  
 Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Boloitin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,  
 de-Montigny,J., Dujon B., Durenz,P., Lepligle,A., Llorente,B.,  
 Malperuuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Douvet,M.,  
 Wincker,P. and Weissenbach,C.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies

## REFERENCE

AUTHORS

JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 11152876  
 PUBMED 20584722  
 TITLE 2 (bases 1 to 970)  
 Llorente,B., Valperuuy,A., Blandin,G., Artiguenave,F., Wincker,P.  
 and Dujon,B.  
 Genomic exploration of the hemiascomycetous yeasts: 12.  
 Kluyveromyces marxianus var. marxianus

## REFERENCE

AUTHORS

JOURNAL FEMS Lett. 487 (1), 71-75 (2000)  
 MEDLINE 11152887  
 PUBMED 20584722  
 TITLE 3 (bases 1 to 970)  
 Genoscope.  
 Direct Submission.  
 Submitted (08-SEP-2003) Genoscope - Centre National de Séquençage,  
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seque@genoscope.cns.fr - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

## COMMENT

This GSS is part of a random genomic sequencing program of thirteen

yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*  
*exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

## FEATURES

source

Location/Qualifiers  
 1..970  
 /organism="Kluyveromyces marxianus"  
 /mol\_type="genomic DNA"  
 /strain="CBS 712"  
 /variety="marxianus"  
 /db\_xref="taxon:4911"  
 /clone="AZ0AA314C12"  
 /clone\_1lb="AZ0AA"  
 /note="end : 77"  
 278 a 196 c 244 g 251 t 1 others

## BASE COUNT

278 a 196 c 244 g 251 t 1 others

## ORIGIN

Query Match 7.2% Score 19; DB 29; Length 970;  
 Best Local Similarity 100.0%; Pred. No. 2,9e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

36 GTTCACCAACCCGACGCGC 54

## Db

878 GTTCACCAACCCGACGCGC 860

RESULT 11  
 B2577558 1111 bp DNA linear GSS 17-DEC-2002  
 LOCUS B2577558  
 DEFINITION msb2\_5464.y2 msb Pseudomonas aeruginosa genomic clone msb2\_5464,  
 genomic survey sequence.  
 ACCESSION B2577558  
 VERSION B2577558.1 GI:2712613  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 1111)

## REFERENCE

AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.B., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and O'Brien,M.V.  
 Whole-Genome-Sequence Variation Among Multiple Isolates of  
*Pseudomonas aeruginosa* Library  
 J. Bacteriol. (2002) in press  
 CONTACT: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: [ckraymond@u.washington.edu](mailto:ckraymond@u.washington.edu)  
 Class: shotgun.

## JOURNAL

COMMENT

## FEATURES

source

Location/Qualifiers  
 1..1111  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="msb2\_5464"  
 /clone\_1lb="msb"  
 /note="Environmental isolate, whole genomic shotgun  
 library"  
 242 a 319 c 344 g 199 t 7 others

## BASE COUNT

ORIGIN

Query Match 7.2% Score 19; DB 29; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CGTGTGCTGAGATCACC 144  
 |||||  
 Db 245 CGTGTGCTGAGATCACC 227

RESULT 12  
 AA480437 102 bp mRNA linear EST 15-AUG-1997  
 LOCUS ne70a12.s: NCI\_CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:909598, mRNA  
 DEFINITION sequence.  
 ACCESSION AA480437 GI:2208588  
 VERSION AA480437.1 GI:2208588  
 KEYWORDS EST.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-rt@mail.nih.gov](mailto:cgaps-rt@mail.nih.gov)  
 Tissue Procurement: Lee Helmar, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be  
 found through the T.M.A.C.E. Consortium/UNL at:  
[www.bio.linn.gov/bdip/image/image.html](http://www.bio.linn.gov/bdip/image/image.html)  
 Seq primer: 41ml3 fwd. ET from Amerstam  
 High quality sequence stop: 77.

FEATURES  
 source  
 1..102  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:909598"  
 /issue\_type="Ewing's sarcoma"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Ew1"  
 /note="Vector: pAMP10; mRNA made from Ewing's sarcoma.  
 cDNA made by oligo-dT priming. Non-directionally cloned.  
 Size selected on agarose gel; average insert size 600 bp.  
 Reference: Krizman et al. (1996) Cancer Research  
 56:5380-5383."

BASE COUNT 23 a 30 c 26 g 23 t

ORIGIN

Query Match 6.8% Score 18; DB 9; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAGATCACC 144  
 |||||  
 Db 53 GTGCTGTGAGATCACC 36

RESULT 13  
 AA706470 140 bp mRNA linear EST 24-DEC-1997  
 LOCUS ag95c07.s1 StrataGene hnt neuron (#937233) Homo sapiens cDNA clone  
 DEFINITION IMAGE:1142220 3', mRNA sequence.  
 ACCESSION AA706470  
 VERSION AA706470.1 GI:2716388

KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-rt@mail.nih.gov](mailto:cgaps-rt@mail.nih.gov)  
 Tissue Procurement: Lee Helmar, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be  
 found through the T.M.A.C.E. Consortium/UNL at:  
[www.bio.linn.gov/bdip/image/image.html](http://www.bio.linn.gov/bdip/image/image.html)  
 Seq primer: 40ml3 fwd. ET from Amerstam  
 High quality sequence stop: 134.

FEATURES  
 source  
 1..140  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1142220"  
 /issue\_type="HNT neurons"  
 /lab\_host="SCL (Kanamycin resistant)"  
 /clone\_lib="Stratagene hnt neuron (#937233)"  
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dT.  
 Differentiated, post mitotic hnt neurons. Average insert  
 size: 1.5 kb; hnt-ZAP XR Vector; -5' adaptor sequence: 5'  
 CATTGCGACGACG 3' -3' adaptor sequence: 5'  
 CTCAGTTTCTTTTCTTTT 3'."

BASE COUNT 32 a 41 c 33 g 34 t

ORIGIN

Query Match 6.8% Score 18; DB 9; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAGATCACC 144  
 |||||  
 Db 53 GTGCTGTGAGATCACC 36

RESULT 14  
 AV940990 141 bp mRNA linear EST 18-JAN-2002  
 LOCUS AV940990 X. Sato unpublished cDNA library, strain H602 adult,  
 DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum  
 cDNA clone bar30k10 5', mRNA sequence.  
 ACCESSION AV940990  
 VERSION AV940990.1 GI:18236787  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. spontaneum  
 ORGANISM Hordeum vulgare subsp. spontaneum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae  
 / Triticeae; Hordeum.  
 1 (bases 1 to 141)  
 Reference Sato, K., Saito, D. and Takega, K.  
 Barley EST sequencing project in NIG and Okayama Univ  
 Unpublished  
 Contact: Tadashi Shin-i  
 Center for Genetic Resource Information  
 National Institute of Genetics  
 111-1 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855

FEATURES Email: tshin@genes.nig.ac.jp.  
Location: Qualifiers

1.141

/organism="Hordium vulgare subsp. spontaneum"

/mol\_type="mRNA"

/strain="H602"

/db\_xref="taxon:77009"

/clone="hah30x10"

/tissue\_type="top three leaves"

/dev\_stage="adult, heading stage"

/clone\_lib="X. Sato unpublished cDNA library, strain H602"

adult, heading stage top three leaves"

BASE COUNT 46 a 31 c 44 g 18 t 2 others

ORIGIN

Query Match 6.8%; Score 18; DB 9; Length 141;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 AGCAGCTTCAGGTCGCGC 213

DB 68 AGCAGCTTCAGGTCGCGC 85

RESULT 15

LOCUS BQ354983

DEFINITION RC4-HT1111-22:200-022-ec7 HT1111 Homo sapiens cDNA, mRNA sequence.

ACCESSION BQ354983

VERSION BQ354983.1

KEYWORDS

EST

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/getcdna.pl?lib=RC4&c2=RC4-HT1111-

221200-022-e07&c3=2000-12-22&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 60

High quality sequence stop: 96.

Location/Qualifiers

1.192

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="HT1111"

/note="Organ: head neck; Vector: puc19; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (O.S. Lenters Patent application

profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions.

BASE COUNT 45 a 62 c 53 g 32 t

ORIGIN

Query Match 6.8%; Score 18; DB 13; Length 192;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 TGTGCTACGTGAGACTGG 112

DB 35 TGTGCTACGTGAGACTGG 52

Search completed: November 7, 2003, 14:47:14  
Job time: 1320 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM: nucleic - nucleic search, using sw model

```
Run on: November 7, 2003, 12:57:06 / Search time 188 seconds
        'without alignments'
        3790.700 Million cells updates/second
```

US-09-622-5003-3

Sequence: 1 atgatcaaggtggagatcaa.....gacctggcgccccaagtcaa 264

Scoring table: OLIGO\_NUC

Searched: 2552756 seqs, 1349719017 residues

word size :

Total number of hits satisfying chosen parameters: 5:055:2

```

Minimum DB seq length: 3
Maximum DB seq length: 300000000

```

Maximum D3 seq. length: 20000000000

Post-processing: Listing first 45 summaries

Database :

1	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1960.DAT
2	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1961.DAT
3	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1962.DAT
4	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1963.DAT
5	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1964.DAT
6	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1965.DAT
7	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1966.DAT
8	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1967.DAT
9	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1968.DAT
10	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1969.DAT
11	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1970.DAT
12	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1971.DAT
13	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1972.DAT
14	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1973.DAT
15	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1974.DAT
16	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1975.DAT
17	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1976.DAT
18	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1977.DAT
19	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1978.DAT
20	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA1979.DAT
21	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA2000.DAT
22	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA2001.DAT
23	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA2001B.DAT
24	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA2002.DAT
25	/SIDSL/gcgdata/geneesq/geneesqr-emb1/NA2003.DAT

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution..

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	264	100.0	264	2	AAZ11645	Colicapeg W.3 gene
2	19	7.2	3596	23	ABU15453	Drosophila melanog
3	15	7.2	5892	23	ABU12454	Drosophila melanog
4	13	7.2	14756	23	AAIC4444	Human yescdactiv
5	13	7.2	25542	3	ABU15452	Drosophila melanog
6	18	6.8	487	2	AAIC7959	Arabidopsis thalian
7	18	6.8	586	2	AAIC6777	Arabidopsis thalian
8	18	6.8	587	22	AAH13570	Human cDNA clone

1	18	6.9	610	24	AB086303	Arabidopsis thaliana
2	18	6.9	1142	25	AC050493	Human secreted pro
3	18	6.9	1142	25	AB271271	Human secreted pro
4	18	6.8	1265	24	ABK53312	Human CDNA encodin
5	18	6.8	1780	20	AA093114	DNA encoding N. me
6	14	6.8	1782	21	AA014446	N. meningitidis Me
7	18	6.8	1927	22	AA014055	Human cDNA sequenc
8	19	6.8	2078	22	AA016785	Human cDNA sequenc
9	18	6.8	2079	24	AA024731	Human secreted pro
10	18	6.8	2433	21	AA073265	Human ORF05P2400
11	18	6.8	3951	22	AA011977	Mycobacterium tuberc
12	18	6.8	4063	22	AB017059	Mycobacterium me,ano
13	18	6.8	5022	20	AA008417	Human strain AD169
14	18	6.8	7242	20	AA008411	Human strain AD169
15	18	6.9	7920	21	AA023880	Mycobacterium me,ano
16	18	6.9	49914	21	AA041471	N. meningitidis pa
17	18	6.8	323554	24	AB074113	Human cytochrome P450
18	18	6.8	346980	21	AA011677	Mycobacterium me,ano
19	18	6.8	346980	21	AB044649	Mycobacterium me,ano
20	18	6.8	346980	21	AA041690	N. meningitidis B
21	18	6.8	143166	21	AA019663	Mycobacterium tuberc
22	18	6.8	445376	22	AA019663	Mycobacterium tuberc
23	18	6.8	445376	22	AA019663	Mycobacterium tuberc
24	18	6.8	445376	22	AA019663	Mycobacterium tuberc
25	18	6.8	445376	22	AA019663	Mycobacterium tuberc
26	18	6.8	445376	22	AA019663	Mycobacterium tuberc
27	18	6.8	445376	22	AA019663	Mycobacterium tuberc
28	18	6.8	445376	22	AA019663	Mycobacterium tuberc
29	18	6.8	445376	22	AA019663	Mycobacterium tuberc
30	18	6.8	445376	22	AA019663	Mycobacterium tuberc
31	18	6.8	445376	22	AA019663	Mycobacterium tuberc
32	18	6.8	445376	22	AA019663	Mycobacterium tuberc
33	18	6.8	445376	22	AA019663	Mycobacterium tuberc
34	18	6.8	445376	22	AA019663	Mycobacterium tuberc
35	18	6.8	445376	22	AA019663	Mycobacterium tuberc
36	18	6.8	445376	22	AA019663	Mycobacterium tuberc
37	18	6.8	445376	22	AA019663	Mycobacterium tuberc
38	18	6.8	445376	22	AA019663	Mycobacterium tuberc
39	18	6.8	445376	22	AA019663	Mycobacterium tuberc
40	18	6.8	445376	22	AA019663	Mycobacterium tuberc
41	18	6.8	445376	22	AA019663	Mycobacterium tuberc
42	18	6.8	445376	22	AA019663	Mycobacterium tuberc
43	18	6.8	445376	22	AA019663	Mycobacterium tuberc
44	18	6.8	445376	22	AA019663	Mycobacterium tuberc
45	18	6.8	445376	22	AA019663	Mycobacterium tuberc

## ALIGNMENTS

## RESULT

AAZI1645 standard; DNA; 264 BP

AC AAZ-2645;

DT 13-NCV-1993 (first entry)

DE Coliphage M13 gene 5 synthetic nucleic acid sequence

KW' Geminivirus resistance; ssDNA virus; ssDNA-binding protein; IVV

XV plant virus; plant protection; ss.

OS Coliphage sp

**X**

FT  
CDS $\times \times$  :

XX :  
:  
:  
:  
:  
:  
:  
:  
:  
:

XX 0  
2  
0  
1  
2  
2  
2  
2

$\times$

[illegible]

XX  
XX



PA	(PEKE ) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EM;
DR	WPI: 2001-656860/75.
DR	P-FSDB: ABB68351.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signaling and cell-cell
XX	interactions .
PS	
XX	Claim 1: SEQ ID NO 31844; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signaling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB131840-AB131851), expressed DNA
CC	sequences (AB131840-AB131851) and the encoded proteins
CC	(AB85737-AB872372).
CC	
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pcf_sequences.
XX	
XX	Sequence 5892 BP; 1870 A; 983 C; 1009 G; 2030 T; 0 other;
XX	
Query Match	7.2%; Score 19; DS 23; Length 5892;
Best Local Similarity	100.0%; Pred. No. 31;
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	
	78 CAGCCTGAACGAGCAGCACTG 96
	614 CAGCCTGAACGAGCAGCACTG 596
Db	
RESULT 4	
AAL04404	
ID	AAL04404 standard; DNA; 14769 BP.
XX	
XX	AAL04404;
XX	
DT	21-NOV-2001 (first entry)
XX	
DS	Human reproductive system related antigen DNA SEQ ID NO: 7092.
XX	
XX	Human; reproductive system related antigen; reproductive system disorder;
XX	cancer; gene therapy; ds
XX	
CS	Homo sapiens.
XX	
XX	
PN	WO200155320-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01339.
XX	
XX	
PR	31-JAN-2003; 2000US-0179065.
PR	04-FEB-2003; 2000US-0180628.
PR	24-FEB-2003; 2000US-0184664.
PR	02-MAR-2003; 2000US-0186350.
PR	16-MAR-2003; 2000US-0189974.
PR	17-MAR-2003; 2000US-0190376.
PR	18-APR-2003; 2000US-0198123.
PR	19-MAY-2003; 2000US-020515.
PR	07-JUN-2003; 2000US-0209467.
PR	28-JUN-2003; 2000US-0214086.
PR	30-JUN-2003; 2000US-0215135.
PR	07-JUL-2003; 2000US-0216647.
PR	07-JUL-2003; 2000US-021688C.
PR	11-JUL-2003; 2000US-0217487.
PR	11-JUL-2003; 2000US-0217496.
PR	14-JUL-2003; 2000US-021829C.





PR 01-SEP-1993; 99US-015193C.  
PR 07-SEP-1993; 99US-0152363.  
PR 10-SEP-1993; 99US-0153073.  
PR 13-SEP-1993; 99US-0153759.  
PR 15-SEP-1993; 99US-0154018.  
PR 16-SEP-1993; 99US-0154039.  
PR 20-SEP-1993; 99US-0154719.  
PR 22-SEP-1993; 99US-0155119.  
PR 23-SEP-1993; 99US-0155486.  
PR 24-SEP-1993; 99US-0156539.  
PR 28-SEP-1993; 99US-0156458.  
PR 29-SEP-1993; 99US-0156596.  
PR 04-OCT-1993; 99US-0157117.  
PR 05-OCT-1993; 99US-0157753.  
PR 06-OCT-1993; 99US-0157865.  
PR 07-OCT-1993; 99US-0158232.  
PR 08-OCT-1993; 99US-0158232.  
PR 12-OCT-1993; 99US-0158369.  
PR 13-OCT-1993; 99US-0159293.  
PR 13-OCT-1993; 99US-0159294.  
PR 14-OCT-1993; 99US-0159295.  
PR 14-OCT-1993; 99US-0159329.  
PR 14-OCT-1993; 99US-0159330.  
PR 14-OCT-1993; 99US-0159331.  
PR 14-OCT-1993; 99US-0159637.  
PR 14-OCT-1993; 99US-0159638.  
PR 18-OCT-1993; 99US-0159584.  
PR 21-OCT-1993; 99US-0160741.  
PR 21-OCT-1993; 99US-0160767.  
PR 21-OCT-1993; 99US-0160768.  
PR 21-OCT-1993; 99US-0160770.  
PR 21-OCT-1993; 99US-0160814.  
PR 21-OCT-1993; 99US-0160815.  
PR 22-OCT-1993; 99US-0160980.  
PR 22-OCT-1993; 99US-0160981.  
PR 22-OCT-1993; 99US-0160989.  
PR 23-OCT-1993; 99US-0161404.  
PR 23-OCT-1993; 99US-0161405.  
PR 25-OCT-1993; 99US-0161406.  
PR 26-OCT-1993; 99US-0161359.  
PR 26-OCT-1993; 99US-0161360.  
PR 26-OCT-1993; 99US-0161361.  
PR 28-OCT-1993; 99US-0161942.  
PR 28-OCT-1993; 99US-0161993.  
PR 28-OCT-1993; 99US-0161993.  
PR 29-OCT-1993; 99US-0162142.

Query March 6.88; Score 18; DB 21; Length 487;  
Best Local Similarity 100.0%; Pred. No. 1e-02; 0; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 177 CCGTACACCGTGACCT 194  
DB 93 CCGTACACCGTGACCT 110

RESULT 7  
AAC46777

ID AAC46777 standard; DNA; 586 BP.

AC AAC46777;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 51363.

KM Hybridisation assay; genetic mapping; gene expression control;

KW Protein identification; signal transduction pathway;

XX Metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EPI033405-A2.

XX 06-SEP-2000.  
PR 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121625.  
XX 05-MAR-1999; 99US-0121180.  
XX 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128237.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132046.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134931.  
PR 21-MAY-1999; 99US-0135194.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135829.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137628.  
PR 04-JUN-1999; 99US-0137802.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138240.  
PR 13-JUN-1999; 99US-0138567.  
PR 14-JUN-1999; 99US-0139219.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139482.  
PR 18-JUN-1999; 99US-0139484.  
PR 18-JUN-1999; 99US-0139485.  
PR 18-JUN-1999; 99US-0139486.  
PR 18-JUN-1999; 99US-0139487.  
PR 18-JUN-1999; 99US-0139488.  
PR 19-JUN-1999; 99US-0139489.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140635.  
PR 26-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.

```

PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144632.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145086.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 28-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 04-AUG-1999: 99US-0147036.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149829.
PR 22-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150565.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151303.
PR 01-SEP-1999: 99US-0151303.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 13-SEP-1999: 99US-0154718.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.

```

```

PR 22-SEP-1999: 99US-0155119.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 26-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156536.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157866.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158169.
PR 13-OCT-1999: 99US-0158293.
PR 13-OCT-1999: 99US-0158294.
PR 13-OCT-1999: 99US-0158294.
PR 14-OCT-1999: 99US-0158329.
PR 14-OCT-1999: 99US-0158330.
PR 14-OCT-1999: 99US-0158331.
PR 14-OCT-1999: 99US-0158337.
PR 14-OCT-1999: 99US-0158636.
PR 19-OCT-1999: 99US-0158584.
PR 21-OCT-1999: 99US-0160747.
PR 21-OCT-1999: 99US-0160767.
PR 22-OCT-1999: 99US-0160768.
PR 22-OCT-1999: 99US-0160770.
PR 22-OCT-1999: 99US-0160814.
PR 22-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160983.
PR 23-OCT-1999: 99US-0161454.
PR 25-OCT-1999: 99US-0161455.
PR 25-OCT-1999: 99US-0161456.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 26-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

```

Query Match Best Local Similarity 6.6% Score 18: DB 21: Length 586:  
Matches 18: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

CY 177 CCTGTACACCGTGCACCT 154
    |||||
Db 149 CCTGTACACCGTGCACCT 156

```

```

RESULT #
AAH13510/c
ID AAH13510 standard: cDNA, 587 BP.
AC AAH13510;
XX
XX 26-SEP-2001 (first entry:
DT
DE Human cDNA clone (3' primer) SEQ ID NO:10365.
XX
XX Human (primer) detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EN EP107467-A2.
XX
PD 07-FEB-2001.
PF
PF 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999: 99US-0248036.
XX 27-AUG-1999: 99US-0300253.
XX 11-JAN-2000: 2000JP-018776.
XX 02-MAY-2000: 2000JP-0183767.

```

```

PR 09-JUN-2000; 2200CP-224:899.
XX (HELI-) HELIX RES INST.
XX
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-3:8749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Claim 3: SEQ ID 10365; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialized methods. AAH3166 to AAH1628 and
XX AAH1633 to AAH18742 represent human cDNA sequences; AAB9446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
SQ Sequence 587 BP; 137 A; 152 C; 126 G; 165 T; 7 other;
XX
XX Query Match 6.8%; Score 18; DB 22; Length 587;
XX Best Local Similarity 100.0%; Pred. No. 1e-02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 127 GTGCTGCTGAGATCACC 144
XX |||||
XX 295 GTGCTGCTGAGATCACC 276
XX
XX RESULT 9
XX ABQ66303/c
XX ID ABQ66303 standard; DNA; 610 BP.
XX
XX AC ABQ66303.
XX
XX D7 21-AUG-2002 (first entry);
XX
XX DE Arabidopsis thaliana polynucleotide SEQ ID NO 880.
XX
XX KM Arabidopsis thaliana thale cress; plant; transgenic; GYO; disease;
XX stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;
XX insecticide; antibiotic; ds.
XX
XX KM Arabidopsis thaliana.
XX
XX OS
XX PN US2002059663-A1.
XX
XX PD 16-MAY-2002.
XX
XX XX 26-JAN-2002; 2001US-0770149.
XX
XX PF
XX XX 27-JAN-2000; 2000US-178506P.

```

```

XX (GRL) GORLACH J.
XX (ANY) AN Y.
XX (HMT) HAMILTON C M.
XX (PRC) PRICE J L.
XX (RAIN) RAINES T M.
XX (YUY) YU Y.
XX (RAME) RAMEKA J G.
XX (PAGE) PAGE A.
XX (MATH) MATHEW A V.
XX (LEDF) LEDFORD B L.
XX (WOES) WOESSNER J P.
XX (HAAS) HAAS W D.
XX (GARC) GARCIA C A.
XX (KRIC) KRICKER M.
XX (SLAT) SLATER T.
XX (DAVI) DAVIS K R.
XX (ALEN) ALLEN K.
XX (HOFF) HOFFMAN N.
XX (HURB) HURBAN P.
XX
XX 361a J. An Y, Hamilton CM, Price JL, Raines TM, Yu Y,
XX Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,
XX Hurban P;
XX WPI; 2002-479224/51.
XX
XX New nucleic acid that hybridizes to Arabidopsis thaliana sequences,
XX useful e.g. for preparing transgenic plants with increased resistance
XX or altered metabolism -
XX
PS Claim 1; SEQ ID NO 880; 40pp + Sequence Listing; English.
XX
XX The invention relates to nucleic acids (i) that hybridize under stringent
XX conditions to any of 399 sequences (ABQ65424-ABQ66422) or their
XX fragments; (i) are used to express the corresponding polypeptides (ii) or
XX to produce genetically modified plant cells or transgenic plants, which
XX may have improved resistance to disease or stress, or altered
XX metabolic/biosynthetic pathways (for production of commercial
XX nutritional or medicinal products), or generally any trait of interest,
XX or can be used to screen for biologically active agents (e.g. fungicides,
XX insecticides and antibiotics).
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO at seqdata.uspto.gov/sequence.html?docid=93939770149.
XX
XX Sequence 610 BP; 175 A; 152 C; 151 G; 130 T; 0 other;
XX
XX Query Match 6.8%; Score 18; DB 24; Length 610;
XX Best Local Similarity 100.0%; Pred. No. 1e-02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 177 CCTGTACACCGTCCACT 194
XX |||||
XX 487 CCTGTACACCGTCCACT 470
XX
XX RESULT 10
XX ACC50489
XX ID ACC50489 standard; cDNA; 1142 BP.
XX
XX AC ACC50489.
XX
XX D7 12-JUN-2003 (first entry);
XX
XX DE Human secreted protein coding sequence, SEQ ID 156.
XX
XX KM Cardiac; antiarrhythmic; antiarteriosclerotic; vasotrophic; cytostatic;
XX vulnerable; antiinflammatory; neurotrophic; neuroprotective;
XX antiparkinsonian; gene therapy; human; cardiovascular disorder;
XX gene; ss.
XX

```





KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;  
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;  
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;  
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;  
 KW tissue regeneration; wound healing; burn; haematopoiesis;  
 KW myeloid cell deficiency; lymphoid cell deficiency.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200177288-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 29-MAR-2001; 2001WO-US10224.  
 XX  
 PR 06-APR-2000; 2000US-195582P.  
 XX  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick NJ,  
 PI Gullotta K, Graham JR;  
 XX  
 DR WPI; 2002-179321/23.  
 XX  
 PT Five hundred and ninety two polynucleotides derived from a variety of  
 PT human tissue sources which encode secreted proteins, useful for  
 PT treating immune deficiencies and disorders such as autoimmune disorders  
 PT  
 XX  
 PS Claim 1; Page 299; 372pp; English.  
 XX  
 CC The invention relates to 592 polynucleotides which have been derived from  
 CC a variety of human tissue sources and which encode novel secreted  
 CC proteins. The polynucleotides can be used as probes for the  
 CC identification and isolation of full length cDNA and genomic DNA. The  
 CC polynucleotides and proteins can also be used as nutritional supplements.  
 CC The proteins are useful in the treatment of various immune deficiencies.  
 CC and disorders such as viral infections, bacterial infections, fungal  
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple  
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions  
 CC and conditions (e.g. asthma). They are also useful for treating  
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's  
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),  
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also  
 CC useful for tissue regeneration, for wound healing and in the treatment of  
 CC burns, incisions and ulcers. The proteins are also useful for regulating  
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.  
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.  
 CC  
 XX  
 SQ Sequence 1265 BP; 368 A; 278 C; 308 G; 311 T; 0 other;  
 Query Match 6.8%; Score 18; DB 24; Length 1265;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 127 GTGCTGTGAAGATCACC 144  
 |||||  
 Db 971 GTGCTGTGAAGATCACC 986  
 |||||

OS Neisseria meningitidis.  
 XX  
 PN WO9936344-A2.  
 XX  
 PD 22-JUL-1999.  
 XX  
 PF 14-JAN-1999; 99WO-IB00103.  
 XX  
 PR 09-OCT-1998; 98GB-0022143.  
 XX  
 PR 14-JAN-1998; 98GB-0000760.  
 PR 01-SEP-1998; 98GB-0019015.  
 XX  
 PA (CHIRON) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizzo M, Rappuoli R, Scarlato V;  
 XX  
 DR WPI; 1999-444400/37.  
 XX  
 DR P-2528; AAY27232.  
 XX  
 PT New protein and its nucleotide sequence, useful in vaccines or  
 PT diagnostic compositions for treating and/or preventing Neisseria  
 PT meningitidis infections  
 PT  
 XX  
 PS Claim 9; Page 95; 123pp; English.  
 XX  
 CC The invention provides proteins (AAY27232-245) from Neisseria  
 CC meningitidis (strains A and B) and nucleic acid sequences (AAK93123-167)  
 CC encoding the proteins. Compositions comprising the protein, nucleic acid  
 CC or antibody specific to the protein are useful as pharmaceuticals, e.g. a  
 CC vaccine composition or a diagnostic composition. The composition is also  
 CC useful for treating or preventing an infection due to Neisseria  
 CC bacteria, especially Neisseria meningitidis.  
 CC  
 XX  
 SQ Sequence 1760 BP; 577 A; 477 C; 431 G; 285 T; 10 other;  
 Query Match 6.8%; Score 19; DB 20; Length 1760;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 39 CACCACCCGACGCGGCT 56  
 |||||  
 Db 1615 CACCACCCGACGCGGCT 1632  
 |||||

RESULT 14  
 ID AAK81446  
 XX  
 AC AAK81446 standard; DNA; 1782 BP.  
 XX  
 AC AAK81446;  
 XX  
 LT 04-DEC-2000 (first entry);  
 XX  
 DE N. meningitidis MenB polynucleotide sequence GCF number 117.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PR 09-OCT-1998; 98US-0131794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PA (CHIRON) CHIRON CORP.  
 XX  
 PI Frazer CW, Hickey E, Peterson C, Tettelin B, Wenter CT,  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarlato V;  
 XX

PI Rappaport R. Pizsa M.  
 XX  
 DR WPI: 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N. gonorrhoea* -  
 XX  
 PS Disclosure, Page 229, 1760pp: English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AAA81451 to AAA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences. AAA81260 to AAA81333 and AAA82520 to AAA82563 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins. AAA82254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament for in the manufacture of a  
 CC medicament for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus* B; against all serotypes;  
 CC and/or against all pathogenic *Neisseria*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus* B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 S2 Sequence 1782 BP; 577 A; 477 C; 431 G; 285 T; 12 other;  
 XX  
 QY Query Match 6.8%; Score 18; DB 21; Length 1782;  
 Db Best Local Similarity 100.0%; Pred. NO. 96;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 39 CACCACCCGAGCGGCGCT 56  
 Db 1617 CACCACCCGAGCGGCGCT 1634  
 XX  
 RESULT 15  
 AAH14055  
 ID AAH14055 standard; cDNA: 1927 BP.  
 XX  
 AC AAH14055;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:1195.  
 XX  
 KW Human; primer; detection; diagnosis; anti-sense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2003; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 22-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX

PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto G;  
 PI Ishii S, Sugiyama T, Wakatsuki A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318745/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8, SEQ ID 1195; 2517bp - CD ROM: English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5612 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full length  
 CC cDNAs easily without any specialised methods. AAH13166 to AAH13628 and  
 CC AAH13631 to AAH18742 represent human cDNA sequences; AAH24446 to  
 CC AAH25893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 S2 Sequence 1927 BP; 484 A; 453 C; 518 G; 452 T; 0 other;  
 XX  
 QY Query Match 6.8%; Score 18; DB 22; Length 1927;  
 Db Best Local Similarity 100.0%; Pred. NO. 95;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 127 GTGCTGTGAGATGATCACC 144  
 Db 1633 GTGCTGTGAGATGATCACC 1650  
 XX  
 Search completed: November 7, 2003, 13:58:54  
 Job time : 195 secs

GenCore Version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 14:25:19 : Search time 135 Seconds

(without alignments) 4320.766 Million cell updates/sec

Title: US-03-622-500B-3

Sequence: 264

1 atgacaaaggtcggagatca.....ggcctggcccccagatga 264

Scoring table: OLIGO-MTC

Gapop 60.0, Gapext 60.0

Searched: 214354 seqs, 1595478879 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database:

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US05\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 2	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 3	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 4	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 5	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 6	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 7	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 8	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 9	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 10	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 11	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 12	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 13	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 14	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 15	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 16	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 17	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 18	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 19	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 20	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 21	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 22	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 23	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 24	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 25	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 26	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 27	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 28	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 29	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 30	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 31	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 32	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 33	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 34	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 35	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 36	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 37	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 38	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 39	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 40	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 41	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 42	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 43	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 44	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1
C 45	19	7.2	8.5	12 US-10-027-632-157148	Sequence 157148, App1

#### ALIGNMENTS

17	17	6.4	498	14 US-10-027-632-157148	Sequence 157148, App1
18	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
19	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
20	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
21	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
22	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
23	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
24	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
25	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
26	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
27	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
28	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
29	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
30	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
31	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
32	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
33	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
34	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
35	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
36	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
37	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
38	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
39	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
40	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
41	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
42	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
43	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
44	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1
45	17	6.4	519	10 US-09-622-500B-3	Sequence 75, App1

US-10-027-632-157148/c

Sequence 157148, Application US/10027632

Publication No. US20030204375A5

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 109827.129

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/197,483

PRIOR FILING DATE: 2000-04-29

PRIOR APPLICATION NUMBER: US 60/185,018

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,367

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/197,483

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-03

NUMBER OF SEQ ID NOS: 325722

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO: 157148

LENGTH: 815

TYPE: DNA

ORGANISM: Human

US-10-027-632-157148

Query Match: 7.2% Score 19; DB 12; Length 815;

Best Local Similarity: 100.0%; Pred. No. 4.3;

Matches 19; Conservative 0; Mismatches 0; Gaps 0;

123 CCCCGTGGTGGAGATC 141

|||||

```
Db 48 CCCCCTGCTGTGAGATC 30

RESULT 2
US-10-027-632-157148/c
: Sequence 157148, Application US/10027632
: GENERAL INFORMATION:
: APPLICANT: Wang, David G.
: TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
: POLYMORPHISMS IN THE HUMAN GENOME
: FILE REFERENCE: 108827.129
: CURRENT APPLICATION NUMBER: US/10/027,632
: CURRENT FILING DATE: 2002-04-30
: PRIOR APPLICATION NUMBER: US 60/218,006
: PRIOR FILING DATE: 2000-07-12
: PRIOR APPLICATION NUMBER: US 60/198,576
: PRIOR FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: US 60/193,483
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: US 60/185,218
: PRIOR FILING DATE: 2000-02-24
: PRIOR APPLICATION NUMBER: US 60/167,363
: PRIOR FILING DATE: 1999-11-23
: PRIOR APPLICATION NUMBER: US 60/156,358
: PRIOR FILING DATE: 1999-09-28
: PRIOR APPLICATION NUMBER: US 60/146,002
: PRIOR FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 325720
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 157148
: LENGTH: 815
: TYPE: DNA
: ORGANISM: Human
US-10-027-632-157148

Query Match
Best Local Similarity 100.0%; Score 19; DB 11; Length 815;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCCCCTGCTGTGAGATC 141
Db 48 CCCCCTGCTGTGAGATC 30

RESULT 3
US-10-156-761-2399/c
: Sequence 2399, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 2399
: LENGTH: 849
: TYPE: DNA
: ORGANISM: Streptomyces avermitilis
: FEATURE:
: NAME/KEY: CSS
: LOCATION: (1)..(849)
US-10-156-761-2399
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 14; Length 649;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 GGTCGGCACTTCGGCAGC 225
Db 664 GGTCGGCACTTCGGCAGC 646

RESULT 4
US-09-764-891-7092
: Sequence 7092, Application US/09764891
: Publication No. US20030077808A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antipodes
: FILE REFERENCE: PCC06
: CURRENT APPLICATION NUMBER: US/09/764,891
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PAM or file wrapper
: NUMBER OF SEQ ID NOS: 10231
: SOFTWARE: Patentn Ver. 2.0
: SEQ ID NO 7092
: LENGTH: 14769
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-891-7092

Query Match
Best Local Similarity 100.0%; Score 19; DB 11; Length 14769;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCCCCTGCTGTGAGATC 141
Db 8773 CCCCCTGCTGTGAGATC 8791

RESULT 5
US-10-156-761-1
: Sequence 1, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 1
: LENGTH: 849
: TYPE: DNA
: ORGANISM: Streptomyces avermitilis
: FEATURE:
: NAME/KEY: Tisc feature
: LOCATION: (4187/15)
: OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 100.0%; Score 19; DB 14; Length 9025603;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 207 GGTGGGCGAGTTGGCAGC 225  
|||||  
Db 2939921 GGTGGGCGAGTTGGCAGC 2939939

## RESULT 6

US-09-770-149-880/C  
Sequence 880, Application US/09770149  
Patent No. US20020059663A1  
GENERAL INFORMATION:  
APPLICANT: Goriach, Jörn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Kameoka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Kicker, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hubbar, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2024 (PARA-013PPV)  
CURRENT APPLICATION NUMBER: US/09779,149  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,506  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 880  
LENGTH: 610  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-149-880

Query Match: 6.8%; Score 18; DB 9; Length 610;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 177 CCTGTACCGGTGGACCT 194  
|||||  
Db 487 CCTGTACCGGTGGACCT 470

## RESULT 7

US-09-822-849A-439  
Sequence 439, Application US/09822849A  
Patent No. US20020045170A1  
GENERAL INFORMATION:  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Agostino, Michael C.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulkota, Kamaakar  
APPLICANT: Graham, James R.  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6403  
CURRENT APPLICATION NUMBER: US/09/822,849A  
CURRENT FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: 60/195,582  
PRIOR FILING DATE: 2000-04-06

NUMBER OF SEQ ID NOS: 598  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 439  
LENGTH: 1265  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-849A-439

Query Match: 6.8%; Score 18; DB 9; Length 1265;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAGATCACC 144  
|||||  
Db 971 GTGCTGTGAGATCACC 988

## RESULT 8

US-10-156-761-5596/C  
Sequence 5596, Application US/10156761  
Publication No. US20030119016A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HISASHI  
APPLICANT: SHIRA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 5596  
LENGTH: 1437  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11..11437)  
US-10-156-761-5596

Query Match: 6.8%; Score 18; DB 14; Length 1437;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 148 GACGAGGCGAGCCGCC 165  
|||||  
Db 159 GACGAGGCGAGCCGCC 142

## RESULT 9

US-10-198-846-10379  
Sequence 10379, Application US/10198846  
Publication No. US20030099744A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
FILE REFERENCE: WRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18

NUMBER OF SEQ ID NOS: 14084  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 10379  
 LENGTH: 2522  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 25, 29, 2522, 2522  
 OTHER INFORMATION: n = A,T,C or G  
 US-10-198-846-10379

Query Match 6.8% Score 18; DB 14; Length 2522;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGGTGAAGATCACC 144  
 Db 2175 GTGCTGGTGAAGATCACC 2192

RESULT 10  
 US-09-712-363-31  
 Sequence 31, Application US/09712363  
 Patent No. US20020164588A1  
 GENERAL INFORMATION:  
 APPLICANT: Eisenberg, David  
 APPLICANT: Rotstein, Sergio H.  
 APPLICANT: Marcotte, Edward M.  
 TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
 FILE REFERENCE: 07419-032001  
 CURRENT APPLICATION NUMBER: US/09/712.363  
 PRIOR FILING DATE: 2000-11-12  
 PRIOR APPLICATION NUMBER: PCT/US00/02246  
 PRIOR FILING DATE: 2000-01-28  
 PRIOR APPLICATION NUMBER: 60/179,531  
 PRIOR FILING DATE: 2000-02-01  
 PRIOR APPLICATION NUMBER: 60/117,844  
 PRIOR FILING DATE: 1999-01-29  
 PRIOR APPLICATION NUMBER: 60/118,236  
 PRIOR FILING DATE: 1999-02-01  
 PRIOR APPLICATION NUMBER: 60/126,593  
 PRIOR FILING DATE: 1999-03-26  
 PRIOR APPLICATION NUMBER: 60/134,093  
 PRIOR FILING DATE: 1999-05-14  
 PRIOR APPLICATION NUMBER: 60/134,092  
 PRIOR FILING DATE: 1999-05-14  
 PRIOR APPLICATION NUMBER: 60/165,124  
 PRIOR FILING DATE: 1999-11-12  
 PRIOR APPLICATION NUMBER: 60/165,086  
 PRIOR FILING DATE: 1999-11-12  
 NUMBER OF SEQ ID NOS: 292  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 31  
 LENGTH: 3951  
 TYPE: DNA  
 ORGANISM: Mycobacterium tuberculosis  
 US-09-712-363-31

Query Match 6.8% Score 18; DB 13; Length 3951;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 CCACCCGACGCGGCTGA 58  
 Db 2175 CCACCCGACGCGGCTGA 2185

RESULT 11  
 US-10-234-007-28/c  
 Sequence 28, Application US/10234007

Publication No. US2003032160A1  
 GENERAL INFORMATION:  
 APPLICANT: Hattaboer, Guus  
 APPLICANT: Verhulst, Karine Cornelia  
 APPLICANT: Schouten, Govert Johan  
 APPLICANT: Uytendaele, Alphonsus Gerardus Cornelis Maria  
 APPLICANT: Bout, Abraham  
 TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN A HUMAN CELL  
 FILE REFERENCE: 4038,1US  
 CURRENT APPLICATION NUMBER: US/10/234,007  
 CURRENT FILING DATE: 2002-09-03  
 PRIOR APPLICATION NUMBER: US/09/549,463  
 PRIOR FILING DATE: 2000-04-14  
 PRIOR APPLICATION NUMBER: 05/129,452  
 PRIOR FILING DATE: 1999-04-15  
 NUMBER OF SEQ ID NOS: 32  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO: 28  
 LENGTH: 29  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: oligonucleotide, synthesized sequence, PCR product generate  
 OTHER INFORMATION: primers US-UP and US-IV-DOWN on template pRUT-CKappa  
 US-10-234-007-28

Query Match 6.4% Score 17; DB 14; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAGGTGAGATCAAGC 22  
 Db 26 CAAGGTGAGATCAAGC 10

RESULT 12  
 US-10-010-942B-20/c  
 Sequence 20, Application US/10010942B  
 Publication No. US20030165496A1  
 GENERAL INFORMATION:  
 APPLICANT: Basi, Gurij  
 APPLICANT: Saidanna, Jose  
 APPLICANT: Vednack, Ted  
 TITLE OF INVENTION: HUMANIZED ANTIBODIES THAT RECOGNIZE  
 TITLE OF INVENTION: BETA AMYLOID PEPTIDE  
 FILE REFERENCE: ELN-002  
 CURRENT APPLICATION NUMBER: US/10/010,942B  
 CURRENT FILING DATE: 2002-12-06  
 PRIOR APPLICATION NUMBER: US 60/251,892  
 PRIOR FILING DATE: 2003-12-06  
 NUMBER OF SEQ ID NOS: 63  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 20  
 LENGTH: 142  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: primer  
 US-10-010-942B-20

Query Match 6.4% Score 17; DB 12; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAGGTGAGATCAAGC 22  
 Db 36 CAAGGTGAGATCAAGC 20

RESULT 13  
 US-10-198-846-13944  
 Sequence 13944, Application US/10198846  
 Publication No. US20030099974A1

```

; GENERAL INFORMATION:
; APPLICANT: Xille, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Stehrmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-19
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 13944
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-13944

```

```

Query Match 6.4%; Score 17; DB 14; Length 341;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 6 CAAGTGAGATCAAGCC 22
DB 210 CAAGTGAGATCAAGCC 226

```

```

RESULT 14
US-09-918-995-28426
; Sequence 28426, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-32
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-23
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO: 28426
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-28426

```

```

Query Match 6.4%; Score 17; DB 14; Length 458;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 8 AGGTGGAGATCAAGCC 24
DB 356 AGGTGGAGATCAAGCC 372

```

```

RESULT 15
US-09-919-580-198
; Sequence 198, Application US/09919560
; Patent No. US2002010832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jianchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121,552
; CURRENT APPLICATION NUMBER: US/09/919,560
; CURRENT FILING DATE: 2001-07-30

```

```

; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 198
; LENGTH: 483
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-198

```

```

Query Match 6.4%; Score 17; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 9 AGGTGGAGATCAAGCC 24
DB 329 AGGTGGAGATCAAGCC 345

```

```

Search completed: November 7, 2003, 15:25:46
Job time : 209 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:57:05 ; Search time 1865 Seconds

(without alignments)  
4694,498 Million cell updates/sec

Title: US-09-622-500B-3

Perfect score: 264

Sequence: : agatcaaggtgagatca.....gcctggccgcgaagtaa 264

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 336368 seqs, 1658189874 residues

Word size: 0

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

```

1: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*

```

Result NO.	Score	Query Match	Length	DB ID	Description
1	264	100.0	264	US-09-622-500B-3	Sequence 3, App.1
2	20	7.6	13231	US-10-437-963-7240C	Sequence 74326, A
3	20	7.6	31291	US-09-620-392-11855	Sequence 11855, A
4	20	7.6	54568	US-09-702-134-8795	Sequence 8795, Ap

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES





```
RESULT 9
US-60-170-373-934/C
; Sequence 934, Application US/60170373
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
; FILE REFERENCE: C1000149
; CURRENT APPLICATION NUMBER: US/60/170,373
; CURRENT FILING DATE: 1999-12-13
; NUMBER OF SEQ ID NOS: 4282
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 934
; LENGTH: 364
; TYPE: DNA
; ORGANISM: Human
US-60-170-373-934
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Length 364;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 123 CCCCCTGCTGCTGATGTC 141
Db 166 CCCCCTGCTGCTGATGTC 148
```

```
RESULT 9
US-09-703-708-10893
; Sequence 10893, Application US/09703708
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xanthomonas campestris Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15804)A
; CURRENT APPLICATION NUMBER: US/09/703,708
; CURRENT FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 1999-11-10
; PRIOR FILING DATE: US 60/183,791
; PRIOR APPLICATION NUMBER: 2000-02-22
; NUMBER OF SEQ ID NOS: 18992
; SEQ ID NO 10893
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-09-703-708-10893
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 31; Length 447;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 216 GTTGGAGCGCTGATGTC 234
Db 99 GTTGGAGCGCTGATGTC 117
```

```
RESULT 10
US-60-164-320-10893
; Sequence 10893, Application US/60164320
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xanthomonas campestris Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15804)A
; CURRENT APPLICATION NUMBER: US/60/164,320
; CURRENT FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 18992
; SEQ ID NO 10893
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
```

```
US-60-164 320-10893
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 70; Length 447;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 216 GTTGGAGCGCTGATGTC 234
Db 99 GTTGGAGCGCTGATGTC 117
```

```
RESULT 11
US-60-183-791-10893
; Sequence 10893, Application US/60183791
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xanthomonas campestris Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15804)B
; CURRENT APPLICATION NUMBER: US/60/183,791
; CURRENT FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 18992
; SEQ ID NO 10893
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-60-183-791-10893
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 72; Length 447;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 216 GTTGGAGCGCTGATGTC 234
Db 99 GTTGGAGCGCTGATGTC 117
```

```
RESULT 12
US-09-270-767-11554/C
; Sequence 11554, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326 034
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 11554
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11554
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 18; Length 487;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 138 GATCACCCTGAGAGGCGC 156
Db 328 GATCACCCTGAGAGGCGC 310
```

```
RESULT 13
US-60-252-833-32151
; Sequence 32151, Application US/60252833
; GENERAL INFORMATION:
; APPLICANT: Geyer, Matthew
; APPLICANT: Geyer, Murray R.
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 105292
; CURRENT APPLICATION NUMBER: US/60/252,833
```

```

CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 43535
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 32151
LENGTH: 551
TYPE: DNA
ORGANISM: Bovine
US-60-252-833-32151
```

```

Query Match: 7.2%; Score 19; DB 80; Length 551;
Best Local Similarity: 100.0%; Pred. No. 1,4e+32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

CY 146 TGACGAGGGCCAGCCCGC 164
DB 67 TGACGAGGGCCAGCCCGC 85
```

```

RESULT 14
US-09-796-765-180/C
Sequence 180, Application US/09796765
GENERAL INFORMATION:
APPLICANT: Rosen, et. al.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 57
FILE REFERENCE: PO-57
CURRENT APPLICATION NUMBER: US/09/796,765
CURRENT FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/187,016
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 477
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 180
LENGTH: 572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (19)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (25)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (51)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (70)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (204)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (232)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (314)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (465)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (496)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (509)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
```

```

LOCATION: (552)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (565)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (570)
OTHER INFORMATION: n equals a,t,g, or c
US-09-796-765-180
```

```

Query Match: 7.2%; Score 19; DB 33; Length 572;
Best Local Similarity: 100.0%; Pred. No. 1,4e+32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

CY 25 AGCCAGGCCAGTTCACCA 43
DB 429 AGCCAGGCCAGTTCACCA 411
```

```

RESULT 15
US-09-867-683-180/C
Sequence 180, Application US/09867682
GENERAL INFORMATION:
APPLICANT: Rosen, et. al.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 57
FILE REFERENCE: PO-57C1
CURRENT APPLICATION NUMBER: US/09/867,682
CURRENT FILING DATE: 2001-03-31
PRIOR APPLICATION NUMBER: 60/197,016
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 477
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 180
LENGTH: 572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (19)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (26)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (51)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (53)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (170)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (204)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (232)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (314)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (465)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (496)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
```

Fri Nov 7 15:26:52 2003

us-09-622-500b-3.rtf

Page 6

```

? LOCATION: (509)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: msc_feature
? LOCATION: (552)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: msc_feature
? LOCATION: (565)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: msc_feature
? LOCATION: (570)
? OTHER INFORMATION: n equals a,t,g, or c
US-09-867-682-180

```

Query Match	7.28;	Score 19;	DB 36;	Length 572;
Best Local Similarity	100.0%;	Pred. No. 1.4e+02;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 25 AGCCAGGCCCACTCACC 43  
 |||||:|||||:  
 Db 429 AGCCAGGCCCACTCACC 411

Search completed: November 7, 2003, 15:19:24  
Job time : 1869 secs

Page 6

Page 6

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:59:59 / Search time 158 Seconds  
(without alignments)  
2906.548 Million cell updates/sec

Title: US-09-622-500b-3  
Perfect score: 264  
Sequence: 1 atgacgaagtcggagatcaaa.....gcctggcgcgcgaagtaa 264

Scoring table: OLIGO\_NJC

Gapop 60.0, Gapext 60.0

Searched: 2459958 seqs, 869762408 residues

Word size: 0

Total number of hits satisfying chosen parameters: 49139.6

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Pending\_Patents\_NA\_New.\*  
1: /cgn2\_6/prodata/2/pna/pct\_NEW\_COMB.seq.\*  
2: /cgn2\_6/prodata/2/pna/USC\_NEW\_COMB.seq.\*  
3: /cgn2\_6/prodata/2/pna/USC7\_NEW\_COMB.seq.\*  
4: /cgn2\_6/prodata/2/pna/USC8\_NEW\_COMB.seq.\*  
5: /cgn2\_6/prodata/2/pna/USC9\_NEW\_COMB.seq.\*  
6: /cgn2\_6/prodata/2/pna/US10\_NEW\_COMB.seq.\*  
7: /cgn2\_6/prodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	7.2	201	7	US-60-487-610-77660
2	19	7.2	201	7	US-60-485-450-39979
3	19	7.2	36530	7	US-60-487-610-19850
4	19	7.2	36530	7	US-60-485-450-12225
5	18	6.8	201	7	US-60-495-114-48489
6	18	6.8	261	5	US-09-540-2310-82396
7	18	6.8	261	5	US-10-653-595-88
8	18	6.8	1823	7	US-60-507-511-4755
9	18	6.8	1928	7	US-60-507-511-4735
10	18	6.8	2115	7	US-60-507-511-2061
11	18	6.8	7860	5	US-09-674-546A-1668
12	18	6.8	42203	5	US-12-367-094-53
13	18	6.8	478484	7	US-60-495-114-16468
14	17	6.4	25	5	US-09-956-604D-112.7
15	17	6.4	25	7	US-60-507-511-44668
16	17	6.4	25	7	US-60-507-481-46080
17	17	6.4	110	6	US-10-242-535A-5290
18	17	6.4	201	7	US-60-487-610-66099
19	17	6.4	201	7	US-60-495-114-54072
20	17	6.4	201	7	US-60-500-337-117831
21	17	6.4	201	7	US-60-500-337-117858
22	17	6.4	201	7	US-60-500-337-117873
23	17	6.4	515	7	US-60-507-511-534
24	17	6.4	544	7	US-60-507-481-6553
25	17	6.4	1696	6	US-10-425-114A-14488
26	17	6.4	2043	6	US-10-425-114A-32590

C 27	17	6.4	16244	7	US-60-495-114-16510	Sequence 16510, A
C 28	17	6.4	17596	1	PCT-US03-20691.2	Sequence 2, App1
C 29	17	6.4	45824	7	US-60-500-337-19843	Sequence 19843, A
C 30	17	6.4	54464	7	US-60-495-114-16529	Sequence 16529, A
C 31	17	6.4	73256	6	US-10-322-281-167	Sequence 167, App
C 32	17	6.4	97727	7	US-60-487-610-19748	Sequence 19748, A
C 33	17	6.4	101517	7	US-60-485-450-11977	Sequence 11977, A
C 34	17	6.4	131789	7	US-60-495-114-16225	Sequence 16225, A
C 35	16	6.1	25	5	US-09-956-604D-141215	Sequence 141215, A
C 36	16	6.1	25	5	US-09-954-427A-37-071	Sequence 371071, A
C 37	16	6.1	76	1	PCT-US03-18714-20194	Sequence 2194, A
C 38	16	6.1	157	1	PCT-US03-18714-13778	Sequence 10778, A
C 39	16	6.1	194	7	US-60-507-481-6930	Sequence 9930, App
C 40	16	6.1	201	7	US-60-487-610-40597	Sequence 40597, A
C 41	16	6.1	201	7	US-60-485-450-112	Sequence 112, App
C 42	16	6.1	201	7	US-60-485-114-1125	Sequence 1125, App
C 43	16	6.1	201	7	US-60-485-114-1130	Sequence 1130, App
C 44	16	6.1	201	7	US-60-485-114-1134	Sequence 1134, App
C 45	16	6.1	201	7	US-60-485-114-1139	Sequence 1139, App

## ALIGNMENTS

RESULT 1  
US-60-487-610-77660  
Sequence 77660, Application US/60487610  
GENERAL INFORMATION:  
APPLICANT: CARILL, Michele  
APPLICANT: HUANG, Hongjit  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS.  
FILE REFERENCE: CLO01469  
CURRENT APPLICATION NUMBER: US/60/487,610  
CURRENT FILING DATE: 2003-07-17  
NUMBER OF SEQ ID NOS: 97101  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 77660  
LENGTH: 201  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-487-610-77660

Query Match 7.2% Score 19; Db 7; Length 201;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

27 123 CCGCGTCTGCGGAGATC 141  
28 CCGCGTCTGCGGAGATC 46

RESULT 2  
US-60-485-450-39979  
Sequence 39979, Application US/60485450  
GENERAL INFORMATION:  
APPLICANT: CARILL, Michele  
APPLICANT: CHANG, Sheng-Yung  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C  
TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
FILE REFERENCE: CLO01470  
CURRENT APPLICATION NUMBER: US/60/485,450  
CURRENT FILING DATE: 2003-07-09  
NUMBER OF SEQ ID NOS: 47853  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 39979  
LENGTH: 201  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-485-450-39979





QY 127 GTGCTGTGAGATCACC 144  
Db 179 GTGCTGTGAGATCACC 196

## RESULT 7

US-10-653-595-96  
Sequence 88, Application US/10653595  
GENERAL INFORMATION:  
APPLICANT: Ruben et. al.  
TITLE OF INVENTION: 95 Human secreted proteins  
FILE REFERENCE: P2027P1C1  
CURRENT APPLICATION NUMBER: US/10/653,595  
CURRENT FILING DATE: 2003-09-03  
PRIOR APPLICATION NUMBER: US 09/397945  
PRIOR FILING DATE: 1999-09-17  
PRIOR APPLICATION NUMBER: PCT/US99/05804  
PRIOR FILING DATE: 1999-03-18  
PRIOR APPLICATION NUMBER: 60/078,566  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,576  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,573  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,574  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,579  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/080,314  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080,312  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/078,578  
PRIOR FILING DATE: 1998-03-19  
Remaining Prior Application data removed - See File Wrapper or PAM.  
NUMBER OF SEQ ID NOS: 470  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 88  
LENGTH: 1142  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-653-595-96

Query Match 6.8%; Score 18; DB 6; Length 1142;  
Best Local Similarity 100.0%; Pred. No. 25;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAGATCACC 144  
Db 790 GTGCTGTGAGATCACC 807

## RESULT 8

US-60-507-511-4755  
Sequence 4755, Application US/60507511  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Mounts, William X  
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
FILE REFERENCE: AM 101081  
CURRENT APPLICATION NUMBER: US/60/507,511  
CURRENT FILING DATE: 2003-10-02  
NUMBER OF SEQ ID NOS: 203623  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 4755  
LENGTH: 1823  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-507-511-4755

Query Match 6.8%; Score 18; DB 7; Length 1823;

Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 127 GTGCTGTGAGATCACC 144  
Db 1532 GTGCTGTGAGATCACC 1549

## RESULT 9

US-60-507-511-4135  
Sequence 4135, Application US/60507511  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Mounts, William X  
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATE  
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
FILE REFERENCE: AM 101081  
CURRENT APPLICATION NUMBER: US/60/507,511  
CURRENT FILING DATE: 2003-10-02  
NUMBER OF SEQ ID NOS: 203623  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 4135  
LENGTH: 1928  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-507-511-4135

Query Match 6.8%; Score 18; DB 7; Length 1928;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAGATCACC 144  
Db 1634 GTGCTGTGAGATCACC 1651

## RESULT 10

US-60-507-511-2061  
Sequence 2061, Application US/60507511  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Mounts, William X  
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATE  
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
FILE REFERENCE: AM 101081  
CURRENT APPLICATION NUMBER: US/60/507,511  
CURRENT FILING DATE: 2003-10-02  
NUMBER OF SEQ ID NOS: 203623  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 2061  
LENGTH: 2115  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-507-511-2061

Query Match 6.8%; Score 18; DB 7; Length 2115;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAGATCACC 144  
Db 1812 GTGCTGTGAGATCACC 1829

## RESULT 11

US-09-674-546A-1668  
Sequence 1668, Application US/09674546A  
GENERAL INFORMATION:  
APPLICANT: Institute for Genetic Research  
APPLICANT: Cation Corporation  
TITLE OF INVENTION: Neisseria Meningitidis Antigens and Compositions  
FILE REFERENCE: CHIR-0334  
CURRENT APPLICATION NUMBER: US/09/674,546A

```

CURRENT FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 3264
SOFTWARE: Patent version 3.2
SEQ ID NO 1668
LENGTH: 7800
TYPE: DNA
ORGANISM: Neisseria meningitidis
US-09-674-546A-1668
```

```

Query Match          6.8%; Score 18; DB 5; Length 7800;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

CY      39 CACCACCCGCGAGCGCGT 55
DB      6486 CACCACCCGCGAGCGCGT 6497
```

```

RESULT 12
US-10-367-094-53/C
Sequence 53; Application US/10367094
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc Malandro
TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
FILE REFERENCE: 529452001500
CURRENT APPLICATION NUMBER: US/10367,094
CURRENT FILING DATE: 2003-02-14
NUMBER OF SEQ ID NOS: 203
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53
LENGTH: 42203
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)....(42203);
OTHER INFORMATION: n = A,T,C or G
US-10-367-094-53
```

```

Query Match          6.8%; Score 18; DB 5; Length 42203;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

CY      68 GCAAGCCCTACAGCCTGA 85
DB      27819 GCAAGCCCTACAGCCTGA 27802
```

```

RESULT 13
US-60-495-114-16468/C
Sequence 16468; Application US/60495114
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
TITLE OF INVENTION: POLYMERH-SMS IN NUCLEIC ACID MOLECULES
TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
FILE REFERENCE: CL001480
CURRENT APPLICATION NUMBER: US/60/495,114
CURRENT FILING DATE: 2003-08-15
NUMBER OF SEQ ID NOS: 91238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16468
LENGTH: 478484
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)....(478484)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-495-114-16468
```

```

Query Match          6.8%; Score 15; DB 7; Length 478484;
```

```

Best Local Similarity 100.0%; Pred. No. 19;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

CY      121 TGGTGAAGATCACTGG 148
DB      328591 TGGTGAAGATCACTGG 328674
```

```

RESULT 14
US-09-396-604D-51217
Sequence 51217; Application US/0996604D
GENERAL INFORMATION:
APPLICANT: Miltmann, Michael
TITLE OF INVENTION: Methods of Genetic Analysis of Escherichia Coli
FILE REFERENCE: 3117.1
CURRENT APPLICATION NUMBER: US/09/956,604D
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/234,049
PRIOR FILING DATE: 2000-09-19
NUMBER OF SEQ ID NOS: 14629
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.2
SEQ ID NO 51217
LENGTH: 25
TYPE: DNA
ORGANISM: E. coli
US-09-396-604D-51217
```

```

Query Match          6.4%; Score 17; DB 5; Length 25;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

CY      212 GCCAGTTCGGCAGCTG 228
DB      4 GCCAGTTCGGCAGCTG 26
```

```

RESULT 15
US-60-507-511-44688
Sequence 44688; Application US/60507511
GENERAL INFORMATION:
APPLICANT: Mounts, William M
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATE
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
FILE REFERENCE: AM 101081
CURRENT APPLICATION NUMBER: US/60/507,511
CURRENT FILING DATE: 2003-10-02
NUMBER OF SEQ ID NOS: 209623
SOFTWARE: Patent version 3.2
SEQ ID NO 44688
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapiens
US-60-507-511-44688
```

```

Query Match          6.4%; Score 17; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

CY      8 AGGTGAATCAAGCCG 24
DB      5 AGGTGAATCAAGCCG 21
```

```

Search completed: November 7, 2003, 15:22:11
Job time : 161 secs
```



ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11509

Query Match 7.2% Score 19; DB 4; Length 1872;  
Best Local Similarity 100.0%; Pred. No. 5.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CGTGGTGGTGAAGATCACC 144  
DB 1143 CGTGGTGGTGAAGATCACC 161

## RESULT 3

US-09-252-991A-3380/C  
Sequence 3380 Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 3380  
LENGTH: 1923  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3380

Query Match 7.2% Score 19; DB 4; Length 1923;  
Best Local Similarity 100.0%; Pred. No. 5.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CGTGGTGGTGAAGATCACC 144  
DB 1027 CGTGGTGGTGAAGATCACC 1009

RESULT 4  
US-09-252-991A-11509  
Sequence 11509 Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11509  
LENGTH: 456  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11509

Query Match 6.8% Score 15; DB 4; Length 456;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 GCGTGAATGACGACCGCC 241  
DB 167 GCGTGAATGACGACCGCC 184

RESULT 5  
US-09-252-991A-11732/C  
Sequence 11732 Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11732  
LENGTH: 1059  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11732

Query Match 6.8% Score 18; DB 4; Length 1059;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 GCGTGAATGACGACCGCC 241  
DB 657 GCGTGAATGACGACCGCC 843

RESULT 6  
US-09-252-991A-11984/C  
Sequence 11984 Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11984  
LENGTH: 1111  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11984

Query Match 6.8% Score 18; DB 4; Length 1111;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TCGGACGCTGATGATCG 235  
DB 152 TCGGACGCTGATGATCG 135

RESULT 7  
US-09-252-991A-12028  
Sequence 12028 Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A

```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12028
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-12028

```

```

Query Match          6.8%; Score 18; DB 4; Length 1797;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      218 TCGGAGCCTGATGTCG 235
Db      1652 TCGGAGCCTGATGTCG 1669

```

```

RESULT 9
US-09-252-991A-11628
; Sequence 1628, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.116
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11628
; LENGTH: 2136
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-11628

```

```

Query Match          6.8%; Score 18; DB 4; Length 2136;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      224 GCGTGATGCGAGCGCC 241
Db      453 GCGTGATGTCGACCGCC 470

```

```

RESULT 9
US-09-252-991A-11913/c
; Sequence 11913, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.116
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11913
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-11913

```

```

Query Match          6.8%; Score 18; DB 4; Length 2334;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      218 TCGGAGCCTGATGTCG 235
Db      2248 TCGGAGCCTGATGTCG 2231

```

```

RESULT 10
US-09-252-991A-11999
; Sequence 11999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.116
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11999
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-11999

```

```

Query Match          6.8%; Score 18; DB 4; Length 2436;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      218 TCGGAGCCTGATGTCG 235
Db      1455 TCGGAGCCTGATGTCG 1472

```

```

RESULT 11
US-09-459-586-5
; Sequence 5, Application US/09459586
; Patent No. 5720957
; GENERAL INFORMATION:
; APPLICANT: Jones, Thomas R.
; TITLE OF INVENTION: Identification of a Human
; TITLE OF INVENTION: Cytochrome P-450 Gene Region Involved in Down-Regulation
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: American Cyanamid Company
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Process #1.0, Version #1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/459,586
; FILING DATE: 02-03-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnhard, Elizabeth M.
; REGISTRATION NUMBER: 31,086
; REFERENCE/DOCKET NUMBER: 32,425-01
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 201-683-2158  
TELEFAX: 201-683-4117  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-459-586-5

Query Match 6.8%; Score 18; DB 1; Length 5020;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ACAGTACCCCGGCTGG 133  
|||||  
DB 2593 ACAGTACCCCGGCTGG 2610

RESULT 12  
US-08-282-696-5  
Sequence 5; Application US/08282696  
Patent No. 5846936  
GENERAL INFORMATION:  
APPLICANT: Jones, Thomas R.  
APPLICANT: Campbell, Ann E.  
TITLE OF INVENTION: Identification of a Human  
CYTOMEGALOVIRUS Gene Region Involved in Down-Regulation of  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: United States  
ZIP: 07470-8426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/282,696  
FILING DATE: 23-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,425-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3305  
TELEFAX: 201-831-3246  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-282-696-5

Query Match 6.8%; Score 18; DB 2; Length 5020;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ACAGTACCCCGGCTGG 133  
|||||  
DB 2593 ACAGTACCCCGGCTGG 2610

RESULT 13

US-08-459-586-3  
Sequence 3; Application US/08459586  
Patent No. 5720957  
GENERAL INFORMATION:  
APPLICANT: Jones, Thomas R.  
APPLICANT: Campbell, Ann E.  
TITLE OF INVENTION: Identification of a Human  
CYTOMEGALOVIRUS Gene Region Involved in Down-Regulation  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Campus Drive  
CITY: Parsippany  
STATE: New Jersey  
COUNTRY: United States  
ZIP: 07054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,586  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,425-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-683-2158  
TELEFAX: 201-683-4117  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7242 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-459-586-3

Query Match 6.8%; Score 18; DB 1; Length 7242;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ACAGTACCCCGGCTGG 133  
|||||  
DB 5311 ACAGTACCCCGGCTGG 5328

RESULT 14  
US-08-282-696-3  
Sequence 3; Application US/08282696  
Patent No. 5846936  
GENERAL INFORMATION:  
APPLICANT: Jones, Thomas R.  
APPLICANT: Campbell, Ann E.  
TITLE OF INVENTION: Identification of a Human  
CYTOMEGALOVIRUS Gene Region Involved in Down-Regulation  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: United States  
ZIP: 07470-8426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/282,696  
FILING DATE: 23-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,425-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3305  
TELEFAX: 201-831-3246  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-282-696-5

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,696
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnhard, Elizabeth M.
; REGISTRATION NUMBER: 33,098
; REFERENCE/DOCKET NUMBER: 32,425-00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3246
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7242 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-282-696-3

```

```

Query Match: 6.8%; Score 18; DB 2; Length 7242;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 116 ACAGTACCCCGTCTGG 133
Db 5311 ACAGTACCCCGTCTGG 5328

```

```

RESULT 15
US-09-103-840A-2
; Sequence 2, Application US/09/103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert J.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20037.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

Query Match: 6.8%; Score 18; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 41 CCACCCGACGCGGCTGA 58
Db 767498 CCACCCGACGCGGCTGA 767515

```

Search completed: November 7, 2003, 14:48:16  
 Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw mode.

Run on: November 7, 2003, 12:27:20 (Search time 1599 seconds)

(without alignments)  
7157.151 Million cell updates/sec

Title: US-09-622-500b-3

Perfect score: 264

Sequence: 1 atgacaaagcggagatcaaa.....gcccgtgcccgcgaagtaa 264

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

GenEmbl:  
1: gb\_ba:  
2: gb\_hg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pac:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vl:  
15: em\_ba:  
16: ex\_fun:  
17: em\_hum:  
18: ex\_in:  
19: em\_mu:  
20: em\_om:  
21: em\_ov:  
22: em\_ov:  
23: em\_pat:  
24: em\_ph:  
25: em\_pl:  
26: em\_ro:  
27: em\_sts:  
28: em\_un:  
29: ex\_vl:  
30: em\_hg\_hum:  
31: em\_hg\_inv:  
32: em\_hg\_other:  
33: em\_hg\_mus:  
34: em\_hg\_pun:  
35: em\_hg\_rtd:  
36: em\_hg\_mam:  
37: em\_hg\_vit:  
38: em\_sy:  
39: em\_hgo\_hum:  
40: em\_hgo\_mus:  
41: em\_hgo\_other:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147.2	55.8	6407	7	FICCG
2	147.2	55.8	6407	7	INXPF
3	147.2	55.8	6971	6	AX006032
4	147.2	55.8	6971	6	BD131:575
5	145.6	55.2	6408	7	INPDX
6	145.6	55.2	6408	7	PPDCG
7	145.6	55.2	7055	6	AX006017
8	145.6	55.2	7055	6	BD131:574
9	145.6	55.2	7775	12	AF464139
10	145.6	55.2	7783	6	AX006032
11	145.6	55.2	7783	6	BD131:571
12	145.6	55.2	8233	12	AF362081
13	145.6	55.2	8879	12	AF218733
14	145.6	55.2	9183	12	AF217317
15	145.6	55.2	9183	12	AF218734
16	145.6	55.2	9183	12	AF464138
17	145.6	55.2	9198	12	AF246447
18	145.6	55.2	9206	12	AF218364
19	145.6	55.2	9225	12	AF246446
20	145.6	55.2	9234	12	AF218353
21	145.6	55.2	9252	12	AF246445
22	145.6	55.2	9264	12	AF246446
23	145.6	55.2	9267	12	AF246449
24	145.6	55.2	9270	12	AF246450
25	145.6	55.2	9273	12	AF246448
26	145.6	55.2	9273	12	AF246451
27	145.6	55.2	9273	12	AF246452
28	145.6	55.2	9273	12	AF246453
29	145.6	55.2	9276	12	AF246454
30	145.6	55.2	9279	12	AF246455
31	144	54.5	716	7	FICGS78
32	144	54.5	6437	7	INX13X
33	144	54.5	6867	12	SYNMBE231V
34	144	54.5	6872	12	SYNMBE2342
35	144	54.5	6871	12	SYNMBE2341
36	144	54.5	6940	12	SYNMBE2344
37	144	54.5	7083	6	AR131:047
38	144	54.5	7083	6	AR261:939
39	144	54.5	7196	12	SYNMBE2341V
40	144	54.5	7196	12	SYNMBE2342V
41	144	54.5	7229	12	SYNMBE2343V
42	144	54.5	7229	12	SYNMBE2344V
43	144	54.5	7232	12	SYNMBE2345V
44	144	54.5	7238	12	SYNMBE2346V
45	144	54.5	7244	12	SYNMBE2347V

## ALIGNMENTS

RESULT 1  
FICCG 6407 bp ss-DNA Circular PHG 27 APR 1993  
LOCUS Bacteriophage f1, complete genome.  
DEFINITION  
ACCESSION U02448.1 GI:166201  
VERSION  
KEYWORDS complete genome; origin of replication.  
SOURCE Enterobacteriophage f1  
ORGANISM Enterobacteriophage f1  
REFERENCE 1. (bases 1 to 6407).  
AUTHORS Hill, D.F. and Petersen, G.B.  
TITLE Nucleotide sequence of bacteriophage f1 DNA  
JOURNAL J. Virol. 44 (1), 32-46 (1982)  
MEDLINE 83059882

Pred. No. is the number of results predicted by chance to have a



```

FEATURES             Source
CDS                  6292494
COMMENT              Original
location/Qualifiers  source text: fl DNA from E.coli K08.
                    1. 6407
                    /organism="Enterobacter: a phage fl"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10863"
                    join(6006..6407,2..831)
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02209.1"
                    /db_xref="GI:166202"
                    /translation="MIDMLVRLCEFDISLWCSS:SGNDLAFVLSKATISGMVSA
                    RTVEYHIDGDTVSGLSHPFESLTHYS:AFKYEBSKHFYPCVETIKSPKRVDSH
                    NPGTDLALGSEALUNFASLFCILYDLDAVATTSR:DATFSARAPEN:AKQVI
                    DLRNVSNGQKSTSRSONBETVWNETSRRTVA:YCKVELOHO:QOLSKSPKAM
                    TSYQEOLKTV:SNPDLLEFASGLVRFARFKITRYLKFGLPLDPA:SPAS:YNQOG
                    KD:LFDLMSFSES:LFKAFESDSMNLYDSAV:DAIOSKFTT:TPSKGTSFAKAPYF
                    GFYRLVNEGYSVA:TMPRNSFWRYVALVEGCI:PKSGLMNSTNNVPLVRF:NV
                    DFSGQPDWMEPYLKIA"
                    1..831
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02210.1"
                    /db_xref="GI:166203"
                    /translation="MNIYDSAV:DAIOSKFTT:PSGK:SPKASRYFGFYRYVNE
                    GYDSVAL:MPNSFWRYVALVEGCI:PKSGLMNSTNNVPLVRF:NVDFSGQPDW
                    MEPYLKIA"
                    843..1106
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02211.1"
                    /db_xref="GI:166204"
                    /translation="MKEVEIKPSQACFTTSGVSRQKPYSLNEOLCYVDLGNPPV:
                    VKITDEGAPAYAGLYTVHLSSFKVQCFSLMIDRL:VPAK"
                    1108..1203
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02212.1"
                    /db_xref="GI:166205"
                    /translation="MEQVADSDTLYQMW:QISVVLCPALGI:NGGR"
                    1206..1304
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02213.1"
                    /db_xref="GI:166206"
                    /translation="MSVLVVSFAFVFLGWCLRSGLTYFRLMETSS"
                    1301..1522
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02214.1"
                    /db_xref="GI:166207"
                    /translation="MKKSLV:KASVAVATVPLMSPAEGDDPKAAFPSC:ASAREY
                    IGVAMWVWV:VSAITIGIKL:PKFTSKAS"
                    1579..2853
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02215.1"
                    /db_xref="GI:166208"
                    /translation="MKKCL:PAID:LVVPFSGASV:ESCLAKHTENSFTNKKDXT
                    LDRYANEGCLMNTSVVCTGDETCQYGT:WYDZGLAIFENEGSGSGSGSGSG
                    GGKTPETGDT:PGVYINPLDGTYPG:EQNPANPNP:LESQPLAT:FMFNRR
                    NROGALTIVTGTQDTPVKTIVYQYVPSKAMYAYNMGRFDCAFHSGFEDPFV
                    CEYSGSSDLPQPPVNAAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSDF

```

```

CDS                  6292494
COMMENT              Original
location/Qualifiers  source text: fl DNA from E.coli K08.
                    1. 6407
                    /organism="Enterobacter: a phage fl"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10863"
                    join(6006..6407,2..831)
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02209.1"
                    /db_xref="GI:166202"
                    /translation="MIDMLVRLCEFDISLWCSS:SGNDLAFVLSKATISGMVSA
                    RTVEYHIDGDTVSGLSHPFESLTHYS:AFKYEBSKHFYPCVETIKSPKRVDSH
                    NPGTDLALGSEALUNFASLFCILYDLDAVATTSR:DATFSARAPEN:AKQVI
                    DLRNVSNGQKSTSRSONBETVWNETSRRTVA:YCKVELOHO:QOLSKSPKAM
                    TSYQEOLKTV:SNPDLLEFASGLVRFARFKITRYLKFGLPLDPA:SPAS:YNQOG
                    KD:LFDLMSFSES:LFKAFESDSMNLYDSAV:DAIOSKFTT:TPSKGTSFAKAPYF
                    GFYRLVNEGYSVA:TMPRNSFWRYVALVEGCI:PKSGLMNSTNNVPLVRF:NV
                    DFSGQPDWMEPYLKIA"
                    1..831
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02210.1"
                    /db_xref="GI:166203"
                    /translation="MNIYDSAV:DAIOSKFTT:PSGK:SPKASRYFGFYRYVNE
                    GYDSVAL:MPNSFWRYVALVEGCI:PKSGLMNSTNNVPLVRF:NVDFSGQPDW
                    MEPYLKIA"
                    843..1106
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02211.1"
                    /db_xref="GI:166204"
                    /translation="MKEVEIKPSQACFTTSGVSRQKPYSLNEOLCYVDLGNPPV:
                    VKITDEGAPAYAGLYTVHLSSFKVQCFSLMIDRL:VPAK"
                    1108..1203
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02212.1"
                    /db_xref="GI:166205"
                    /translation="MEQVADSDTLYQMW:QISVVLCPALGI:NGGR"
                    1206..1304
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02213.1"
                    /db_xref="GI:166206"
                    /translation="MSVLVVSFAFVFLGWCLRSGLTYFRLMETSS"
                    1301..1522
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02214.1"
                    /db_xref="GI:166207"
                    /translation="MKKSLV:KASVAVATVPLMSPAEGDDPKAAFPSC:ASAREY
                    IGVAMWVWV:VSAITIGIKL:PKFTSKAS"
                    1579..2853
                    /note="protein 1"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="AA02215.1"
                    /db_xref="GI:166208"
                    /translation="MKKCL:PAID:LVVPFSGASV:ESCLAKHTENSFTNKKDXT
                    LDRYANEGCLMNTSVVCTGDETCQYGT:WYDZGLAIFENEGSGSGSGSGSGSG
                    GGKTPETGDT:PGVYINPLDGTYPG:EQNPANPNP:LESQPLAT:FMFNRR
                    NROGALTIVTGTQDTPVKTIVYQYVPSKAMYAYNMGRFDCAFHSGFEDPFV
                    CEYSGSSDLPQPPVNAAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSDF

```

```

BASE COUNT          1574 a 1298 c 1325 g 2210 t
ORIGIN
Query Match          55.8% Score 147.2; DB 7; Length 6407;
Best Local Similarity 72.3%; Pred. No. 7,9e-14;
Matches 191; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
1 ATGATCAAGTGGAGATCAAGCCAGCGAGCCAGTTCACCAACCGGACGCGCTAGC 60
2 |||||
3 843 ATGATTAAGTTGAATTAACATCTCAAGCGCAATTCACCAACCGCTCTGCTGCTTCT 902
4 |||||
5 61 CGCCAGGCGAGCGCCCTACAGCGCTGACGAGCGAGCGCTGACGAGCGAGCGAG 120
6 |||||
7 903 CGTCAGGCGAGCGCTTCTTACCTGAAAGAGAGAGCTTTGTTACGTTGATTTGATGAA 962
8 |||||
9 121 TAACCGGCTGCTGATTAAGATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
10 |||||
11 963 TATCCGAGCTTCTCAAGATATCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022
12 |||||
13 281 TAACCGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
14 |||||
15 1023 TACACCGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1082
16 |||||
17 241 CTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 284
18 |||||
19 1083 CTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1106
20 |||||

```

```

RESULT 2
INXXF1
LOCUS
DEFINITION Genome of bacteriophage fl (closely related to phages fd and m13).
6407 bp DNA circular PHG 13-FEB-1995

```

```
ACCESSION      V00606 J02449 M10881
VERSION        V00606.1 GI:14974
KEYWORDS       circular; coat protein; genome; origin of replication.
SOURCE         Enterobacteria phage f1
ORGANISM       Enterobacteria phage f1
REFERENCE      1 (bases 1 to 6407);
AUTHORS        Beck, E. and Zink, S.
TITLE          Nucleotide sequence and genome organization of filamentous
JOURNAL        J. Bacteriol. 182:1180-1181 (1990)
MEDLINE        8221180.
PUBMED         6282733
REFERENCE      2 (bases 5493 to 5852)
AUTHORS        La Parina, M. and Vitale, M.
TITLE          Rho-dependence of the terminator active at the end of the 1 region
JOURNAL        Mol. Gen. Genet. 195 (1-2), 5-9 (1984)
MEDLINE        8503586.
PUBMED         6092864
FEATURES       

|        |                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| source | Location/Qualifiers                                                                                                                                                                                                                                                                                                                                                                                                                 |
| gene   | 1..6407                                                                                                                                                                                                                                                                                                                                                                                                                             |
|        | /organism="Enterobacteria phage f1"                                                                                                                                                                                                                                                                                                                                                                                                 |
|        | /mol_type="genomic DNA"                                                                                                                                                                                                                                                                                                                                                                                                             |
|        | /db_xref="taxon:10863"                                                                                                                                                                                                                                                                                                                                                                                                              |
|        | join(6006..6407,1..831)                                                                                                                                                                                                                                                                                                                                                                                                             |
|        | /gene="f1"                                                                                                                                                                                                                                                                                                                                                                                                                          |
|        | join(6006..6407,1..831)                                                                                                                                                                                                                                                                                                                                                                                                             |
|        | /gene="f1"                                                                                                                                                                                                                                                                                                                                                                                                                          |
|        | /codon_start=1                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /transl_table=1                                                                                                                                                                                                                                                                                                                                                                                                                     |
|        | /protein_id="CAA23876.1"                                                                                                                                                                                                                                                                                                                                                                                                            |
|        | /db_xref="GI:603850"                                                                                                                                                                                                                                                                                                                                                                                                                |
|        | /db_xref="SWISS-PROT:P03659"                                                                                                                                                                                                                                                                                                                                                                                                        |
|        | /translation="MIDMLVLRPFLDSVCRSLSGNDLIAVFDKSLATISGMKLSA<br>RTVEHIDGLTVSGLSHPESLPHTSGIARKEGSRNFPQVEYIASFAKVLGGH<br>NVRGTDLALCSALLLNFRANSLPCLYLDJDNATLISLIDATPSARAPNENAKQVI<br>DHLRNSNGQTKSRQSNWSTVTWNETRHRITVAALQVHQLQIQQLSSKSAAM<br>TSYQKQLKVLNPLDLFEASGLVREARLFRVLYKSGFLPLNLDARFASDYNQ3<br>KDIQFDLSPFSELFRFEGSGSNLYDSAVLDAIQSHPTITPSGKTSFAKSRVF<br>GFRLPLNDEGYSVALTMPRNSFMRVYSALVSCGI.FKSQLMNLSTCNVPLVLRFINV<br>DESSORPLWNEPVAKIA" |
|        | 496..831                                                                                                                                                                                                                                                                                                                                                                                                                            |
|        | /note="gene X"                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /codon_start=1                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /trans_table=1                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /protein_id="CAA23867.1"                                                                                                                                                                                                                                                                                                                                                                                                            |
|        | /db_xref="GI:14975"                                                                                                                                                                                                                                                                                                                                                                                                                 |
|        | /db_xref="SWISS-PROT:P03659"                                                                                                                                                                                                                                                                                                                                                                                                        |
|        | /translation="MNYDQSAVLDAIQSHPTITPSGKTSFAKSRVFGRRLAKS<br>GYDSVALTMPRNSFMRVYSALVSCGIPKSQLMNLSTCNVPLVLRFINVDSQSPDW<br>YNEPVAKIA"                                                                                                                                                                                                                                                                                                      |
|        | 843..1106                                                                                                                                                                                                                                                                                                                                                                                                                           |
|        | /note="gene V"                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /codon_start=1                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /trans_table=1                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /protein_id="CAA23868.1"                                                                                                                                                                                                                                                                                                                                                                                                            |
|        | /db_xref="GI:14976"                                                                                                                                                                                                                                                                                                                                                                                                                 |
|        | /db_xref="SWISS-PROT:P03669"                                                                                                                                                                                                                                                                                                                                                                                                        |
|        | /translation="MKEIKPSQAQFTLRSGVSRGSKPYSLNEQLCYVDLQNEYPVL<br>VKITLDEGPAYAPGLYTHLSFKVGGGSLNIDRLRVPAK"                                                                                                                                                                                                                                                                                                                                 |
|        | 1108..1209                                                                                                                                                                                                                                                                                                                                                                                                                          |
|        | /note="gene VII"                                                                                                                                                                                                                                                                                                                                                                                                                    |
|        | /codon_start=1                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /trans_table=1                                                                                                                                                                                                                                                                                                                                                                                                                      |
|        | /protein_id="CAA23869.1"                                                                                                                                                                                                                                                                                                                                                                                                            |
|        | /db_xref="GI:14977"                                                                                                                                                                                                                                                                                                                                                                                                                 |
|        | /db_xref="SWISS-PROT:P03675"                                                                                                                                                                                                                                                                                                                                                                                                        |
|        | /translation="MEGVADFCITIQAMQISVVLCPALGIIAGGQR"                                                                                                                                                                                                                                                                                                                                                                                     |
|        | 1206..1304                                                                                                                                                                                                                                                                                                                                                                                                                          |
|        | /note="gene IX"                                                                                                                                                                                                                                                                                                                                                                                                                     |
|        | /codon_start=1                                                                                                                                                                                                                                                                                                                                                                                                                      |


```

```

/translation="MKEIKPSQAQFTLRSGVSRGSKPYSLNEQLCYVDLQNEYPVL
VKITLDEGPAYAPGLYTHLSFKVGGGSLNIDRLRVPAK"
1108..1209
/note="gene VII"
/codon_start=1
/transl_table=1
/protein_id="CAA23870.1"
/db_xref="GI:14978"
/db_xref="SWISS-PROT:P03677"
/translation="VSVLVYFSAFVGLWCLRSQTYFTFRLETS"
1321..1522
/note="gene VII"
/codon_start=1
/transl_table=1
/protein_id="CAA23871.1"
/db_xref="GI:14979"
/db_xref="SWISS-PROT:P03617"
/translation="MKSLVLAAGVAVTVPMLSFALSGDFPAKAFDSQASATEY
ICNANVYVYVNGATGSLPKFTTSAS"
1579..1853
/note="gene III"
/codon_start=1
/transl_table=1
/protein_id="CAA23872.1"
/db_xref="GI:579103"
/db_xref="SWISS-PROT:P03662"
/translation="MKKLLFAIPVWPVYSHAFVSCAKPTENSFTNWKMDKT
LDVANYEGCLMAVGVVCTGDEFCQGWPIQLAIPVEGGSGSBSGGEGSGS
GOTKPEYGDPIPGYTYINPLDGYTPGTQNPGLANPSELESOPNTFMFKNRER
NKGALTYTGTVGCTGTPVKTVOYTVSSKAMDAYMNGKPRDCAHSGFNDPBY
CEYGGSSDLPPQPVNAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
DYENMANANKAKTENADENLQSDAKGKIDVATDYGAALDGTIDYSGILANNGAT
GPFASNSQMAQVGGSDNSPLMNNFRQYLSPOSVEGRPVFRAGKPYRSTCCDKI
NFRGVFAFLVATFYVFTFANILNKES"
2856..3194
/note="gene VI"
/codon_start=1
/transl_table=1
/protein_id="CAA23873.1"
/db_xref="GI:14981"
/db_xref="SWISS-PROT:P03673"
/translation="PVVLAGPLRLFGFLPVLDQFGLTFJAKGRKIAIASLFL
AAILGNSLVVYLSIDISACHPSPVQGVQVQLPSNALPCYVTVSVAAFIIDVQK
KVSVALDMKX"
3196..4242
/note="gene I"
/codon_start=1
/transl_table=1
/protein_id="CAA23874.1"
/db_xref="GI:14982"
/db_xref="SWISS-PROT:P03657"
/translation="MAVYEVGKLGSEKTLVSGKLCRIYAGCKIATNLDPLQNDP
OGRPAKTRPRVRIIPKPSISDILAIGRNDSDYENKKGSLVLECGKWKTPBMDK
BQPIITFHLARKIKGKIFLVQSLSYDQNSALANRYVCEKRRITLPIVSTL
VSLITGSKPDLKHLGVYKIGSLSTPVEKMTYKSLINATVTRHRSSTDSGV
VSLTFYLSHGRYFKETNLGQKYLITVLEKFSVCLAGFAAFYVSYITPKRE
VKKVQYVDFPDKETLDSGQNLNLYRVFKSKKGLNSDLOXQYSLYIDLCYV
SIKKGNSLIVKCN"
4220..5800
/note="gene IV"
/codon_start=1
/transl_table=1
/protein_id="CAA23875.1"
/db_xref="GI:14983"
/db_xref="SWISS-PROT:P03666"
/translation="MKLVINREVFMLVSSSSFAQVTEMMNSGRCFVTVYSQDTSE
SVYGRDVYGVTVYSSDYKPNLEPFTISVPRANPEMWSISISLCKVAPNNDQVI
DEPPSSNDQEPIDNSAPSGSGFVPPONKAVTQTKINNYRACDLRIYECFKSNTSKS
SVLSDVSGNLLVSAFKQILDNLQFLSTVDLPTDQILIBGLIFEVQSGALDRSPA
ASORGTAVAGVNTDRLTVLSLSAGSGSINGVGLSVALKTNSSKLSLSPRL
LTSQGSGLSVQGNVPFITGRVTVGSATVKNPFCVTEQNCNOISMSVDPVAGGNIV
LDTSKADSLSSTOASDVIITNORSITANNLPDQTLTGLGTYKXTSGSGVFFL
SKPLIGLFLSSRSDSNESTLYIVLVKATIVAA"
1563 a 1296 c 1328 g 2212 c

```

```

BASE COUNT      1563 a 1296 c 1328 g 2212 c
ORIGIN
Query Match:    55.8%; Score 147.2; DB 7; Length 6407;

```



```
/db_xref="GI:9978969"
/translation="MBKKTGYTVDISGMARKHEFEPJSDVACTYKIVQLEDTLE
LKTAKXKXKFPYPAFIHLARLWNSHPFRAMKDELVIDSVPCTVPHQETEF
SLMSVYDDPFLHYSQVACVGEN:AYFPKGFINKMFVSANPWSFTSFLNV
AMNDNFPAPVETMGKYYTGGXVLMPLA:GVHVAVCDGFHGRMLNE:IQVCDEWQGG
A"
6140..6144
6163..6167
6188..6202
/misc_feature
  note="lac operator"
  position=6253..6569
  range=6253..6569
  /note="ompA-FLAG-pep3-gi11s"
  /codon_start=1
  /transl_table=1
  /protein_id="CAC05217.1"
  /db_xref="GI:9528973"
/translation="MKRTAIAIAVALAGFAVYQADYKDVDCIVYHAY:VAKCGGG3
SEFNAGGSGGGSGGSGEGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG
MTNADENALQSPAKGLDVAITDYQAA:DFICVSG:ANGNGATGEGAGSNQMCQ
VCDGNSPLMNNPROY:PSLPSQVEGRPVFGAKKPEFSIDCKIN:FRGVAFLLY
VATFWVFSCTFANILANKES"
6253..6315
/misc_feature
  note="ompA signal sequence"
  position=6316..6327
  range=6316..6327
/misc_feature
  note="FLAG peptide"
  position=6334..6372
  range=6334..6372
/misc_feature
  note="peptide 3"
  position=6394..6569
  range=6394..6569
  /note="gene 111s"
BASE COUNT 1737 a 1402 c 1463 g 2369 t
ORIGIN
Query Match 55.8%; Score 147.2; DB 6; Length 6971;
Best Local Similarity 72.3%; Pred. No. 7,8e-14;
Matches 191; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 1 ATGATCAAGGTGAATCAAGCCAGCCAGCCAGCTTACACCCGCGAGCGGCTGAGC 60
DB 4489 ATGATTAAGTTGAAATTAACCATCTCAAGCGCAATCACTACCGTTGCTTTCT 4548
QY 6 CGCCAGGCGAGGCGCTTACAGCTGAGCGAGCGAGCTGTGCTAGCTGGAGCGAG 120
DB 4549 CGTCAGGCGAGGCGCTTATTCATGAA:TAGACAGCTTTTATGCTTGATTGSGTATATA 4609
QY 121 TACCCCGTGTGTGAAGATCACTCGGACGAGCGAGCGCGCTAGCGCCGCGCTG 180
DB 4609 TATCCGCTGCTTGTCAAGATTACTCTGTATGAGTAGAGTCAAGCGCGCTAGCGCTG 4668
QY 161 TACACGCTGACCTGAGCACTCTCAAGGTGGCGCATTCGCGCAGCGCTGATGACGCGC 240
DB 4669 TACACGCTGACCTGCTCTCTCTGTTCAAGGTGGTCAAGTTGGTTCTTTATGATGACGCT 4728
QY 241 CTGCGCCCTGTGTCGCCGCGCAAGTAA 264
DB 4729 CTGCGCCCTGTGTCGCCGCGCAAGTAA 4752
RESULT 4
BD131875
LOCUS BD131875 6971 bp DNA linear PAT 18-SEP-2002
DEFINITION Novel method and phase for identifying nucleic acid sequence
ACCESSION BD131875
VERSION BD131875.1 GI:23226820
KEYWORDS UP 2002501721-A/7.
SOURCE UP 2002501721-A/7.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 6971)
AUTHORS Rudert, F., Ge, R. and Irig, V.
TITLE Novel method and phase for identifying nucleic acid sequence
JOURNAL Patent: JP 2002501721-A 7 22-JAN-2002;
```

```
MORPHOLOGY AG
COMMENT
CS Artificial Sequence
PN JP 2002501721-A/7
P3 22-JAN-2002
PF 03-AUG-1998 JP 2000505327
PR 01-AUG-1997 EP 9713315.4
PI FRITZ RUDERT, RIMIN GE, VICK IRIG
PC C12N15/09;C12N15/00
CC Description of Artificial Sequence: phage vector (omp3-1B-CC
IRISeg
(circular)
CC gene VI
CC gene I
CC gene IV
CC gene II
CC gene X
CC gene V
CC gene VII
CC gene IX
CC gene VIII
CC cat resistance gene
CC ompA-FLAG-pep3-gi11s
CC ompA signal sequence
CC FLAG peptide
CC peptide 3
CC gene 111s
CC lac operator
CC packaging signal
CC f1 ori
CC f1 terminator
FH Key Location/Qualifiers
FT CDS (94)..(429)
FT CDS (435)..(11478)
FT CDS (1459)..(2736)
FT CDS (3245)..(4474)
FT CDS (4142)..(4474)
FT CDS (4489)..(4749)
FT CDS (4754)..(4852)
FT CDS (4852)..(4959)
FT CDS (4947)..(5165)
FT CDS (5289)..(5945)
FT CDS (6253)..(6569)
FT misc_feature (6253)..(6315)
FT misc_feature (6316)..(6327)
FT misc_feature (6334)..(6372)
FT misc_feature (6394)..(6569)
FT -10 signal (6140)..(6144)
FT -35 signal (6163)..(6167)
FT misc_feature (6188)..(6202)
FT misc_signal (6253)..(6569)
FT rep_origin (3033)..(3249)
FT terminator (491)..(1701)
FEATURES
source
1..6971
Location/Qualifiers
1..6971
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
BASE COUNT 1737 a 1402 c 1463 g 2369 t
ORIGIN
Query Match 55.8%; Score 147.2; DB 6; Length 6971;
Best Local Similarity 72.3%; Pred. No. 7,8e-14;
Matches 191; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 1 ATGATCAAGGTGAATCAAGCCAGCCAGCCAGCTTACACCCGCGAGCGGCTGAGC 60
DB 4489 ATGATTAAGTTGAAATTAACCATCTCAAGCGCAATCACTACCGTTGCTTTCT 4548
QY 6 CGCCAGGCGAGGCGCTTACAGCTGAGCGAGCGAGCTGTGCTAGCTGGAGCGAG 120
DB 4549 CGTCAGGCGAGGCGCTTATTCATGAA:TAGACAGCTTTTATGCTTGATTGSGTATATA 4609
QY 121 TACCCCGTGTGTGAAGATCACTCGGACGAGCGAGCGCGCTAGCGCCGCGCTG 180
```



AGSOGCTVAGGNTVRLTIVSSAGSGFQENGVDVLSVFAKLTNSHKLSTVRLT  
 TSSGKGISIVGQNVFRTGRTGSAVNVNPFCTGROVNGISNVFVPMAGNIV  
 IDITVADSLSSTQASDVITNORSIACTVN-RCSQTLILGLTDYKNTSCDSVPL  
 SKPLIGLFFSRSDSNESSTIYLVKATVIAL"

BASE COUNT 1576 a 1295 c 1325 g 2210 t

ORIGIN

Query Match 55.2%; Score 145.6; DB 7; Length 6408;  
 Best Local Similarity 72.0%; Pred. No. 1,4e-13;  
 Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGATCAAGCTGAGATCAAGCCACGCGCCAGTTTACACACCGGAGCGGCTGAC 60  
 DB 1470 ATGATCAAGCTGAGATCAAGCCACGCGCCAGTTTACACACCGGAGCGGCTGAC 1529  
 QY 61 CGGACGAGGAGCGGCTGACGCTGACGAGCGGCTGACGAGCGGCTGACGAG 120  
 DB 1530 CGGACGAGGAGCGGCTGACGCTGACGAGCGGCTGACGAGCGGCTGACGAG 1589  
 QY 121 TACCGGCTGCTGGTGAAGATCAACCTGAGCGGCGGCGGCTGACGCGGCTG 180  
 DB 1590 TACCGGCTGCTGGTGAAGATCAACCTGAGCGGCGGCGGCTGACGCGGCTG 1649  
 QY 181 TACACGCTGACCTGAGCGGCTGACGAGCGGCTGACGAGCGGCTGACGAGCGG 240  
 DB 1650 TACACGCTGACCTGAGCGGCTGACGAGCGGCTGACGAGCGGCTGACGAGCGG 1705  
 QY 241 CTGCGGCTGCTGGTGAAGATCAACCTGAGCGGCGGCGGCTGACGAGCGG 264  
 DB 1710 CTGCGGCTGCTGGTGAAGATCAACCTGAGCGGCGGCGGCTGACGAGCGG 1733

RESULT 6  
 LOCUS PFDCG 6408 bp ss-DNA circular PHG 28-Apr-1993  
 DEFINITION Bacteriophage fd, strain 478, complete genome  
 ACCESSION J02451 M10731 M10766 M21667 M21668 M21669 M21670 M25198  
 VERSION J02451.1 GI:215394  
 KEYWORDS DNA-binding protein; circular; coat protein; complete genome;  
 origin of replication; unidentified reading frame;  
 Enterobacteria phage fd  
 Enterobacteria phage fd  
 Viruses; ssDNA viruses; Inoviridae; Inovirus.

REFERENCE 1 (sites)  
 Sugimoto, K., Sugisaki, H., Okamoto, T. and Takamami, M.  
 Studies on bacteriophage fd DNA. II. Nucleotide sequence preceding  
 the RNA start-site on a promoter-containing fragment  
 Nucleic Acids Res. 2 (11), 2091-2100 (1975)  
 78248318

REFERENCE 2 (bases 402 to 443)  
 Schaller, H., Gray, C. and Herrmann, K.  
 Nucleotide sequence of an RNA polymerase binding site from the DNA  
 of bacteriophage fd  
 Proc. Natl. Acad. Sci. U.S.A. 72 (2), 737-741 (1975)  
 75139534

REFERENCE 3 (bases 1196 to 1564)  
 Sugimoto, K., Sugisaki, H., Okamoto, T. and Takamami, M.  
 Studies on bacteriophage fd DNA: IV. The sequence of messenger RNA  
 for the major coat protein gene  
 J. Mol. Biol. 110, 487-507 (1977)  
 4 (bases 1 to 6408)

REFERENCE 4 (bases 1 to 6408)  
 Beck, E., Sommer, R., Auerwald, E.A., Kutz, C., Zink, B., Oetelburg, S.,  
 Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takamami, M.  
 Nucleotide sequence of bacteriophage fd DNA  
 Nucleic Acids Res. 5 (12), 4495-4503 (1978)  
 79136480

REFERENCE 5 (bases 5585 to 5771)  
 Gray, C.P., Sommer, R., Polke, C., Beck, E. and Schaller, H.  
 Structure of the origin of DNA replication of bacteriophage fd  
 Proc. Natl. Acad. Sci. U.S.A. 75 (1), 50-53 (1978)  
 745987

REFERENCE 6 (sites)  
 Horabin, J., and Webster, R.E.  
 An amino acid sequence which directs membrane insertion causes loss  
 of membrane potential.  
 J. Biol. Chem. 263 (23), 11575-11583 (1988)  
 88298819

MEDLINE 78116039  
 PUBMED 272686  
 REFERENCE 6 (sites)  
 Horabin, J., and Webster, R.E.  
 An amino acid sequence which directs membrane insertion causes loss  
 of membrane potential.  
 J. Biol. Chem. 263 (23), 11575-11583 (1988)  
 88298819  
 PUBMED 2457024  
 COMMENT Original source text: fd from E.coli.  
 (6) sites: internal signal sequence.  
 (1) sites:  
 single-stranded circular DNA which codes for ten proteins  
 replicative form is duplex, filamentous.  
 (3) missing data project.

## FEATURES

## source

## CDS

1..6408  
 /organism="Enterobacteria phage fd"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10864"  
 join(6007..6408,1..831)  
 /note="11 (x5 replication, nicking)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03303.1"  
 /db\_xref="GI:215395"  
 /translation="MIDMLVLRPFDLSVGRSLSGNDLAFVDSKIATSGNLSA  
 RVEVHIDCLTVSGLSHPFESLPFTYSGIARFVSGSKNFPCEVETKASPAKVLQGH  
 NVGRTDIALGSEALLNPANSPLCTVLDNATIRIRATFAPAPNEMALACVI  
 DILNWSNQRTSTRSQRMESVTFNENSRHRTIYAKHBLQDQIQLSKSPKAKM  
 TSYKEQLVLSNPDLFASJVFKAIEFVLSKFGPLNFDARFASDVNRQS  
 KOLIFLWSFGFSELFKAFESGNIVDQSAVLQSKHPTPSGCTSFKAARVF  
 GYFRRLVWGSDVVALTPWRFNSFMYVALVECGIPKQQLNLSYCNVPLVPRIV  
 DFSQRPDWVNEPVLKIA"  
 496..831  
 /note="X (function unknown)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03304.1"  
 /db\_xref="GI:215396"  
 /translation="KNYEDSAVLEAPVQSKPTTPSGKTSFAKARVFGFRRVNE  
 GDSVALTPRNSFWRYVALVECGIPKQQLNLSYCNVPLVPRIVNDFSSQRPDW  
 VNEPVLKIA"  
 843..1106  
 /note="V (single-stranded binding protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03305.1"  
 /db\_xref="GI:215397"  
 /translation="KKEIKPSQAGTTRGCVSRQCKPVSLNEDLCIVDLNRYVVL  
 VRTLDEQFAYAPGLYTVHSSFKVGGFGSLMIDRLVPAK"  
 1108..1209  
 /note="VII (morphogenesis)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03306.1"  
 /db\_xref="GI:215398"  
 /translation="WEQVACPDITVCAMIQISVLCFALINAGQR"  
 1206..1304  
 /note="IX (minor coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03307.1"  
 /db\_xref="GI:215399"  
 /translation="SVLVYFASFEVLCMLRSGTTFTRIMETSS"  
 <1301..1551  
 /note="mRNA (3' end +/- 14 bp)"  
 1301..1522  
 /note="VIII (major coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03308.1"  
 /db\_xref="GI:215400"

## CDS

1..6408  
 /organism="Enterobacteria phage fd"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10864"  
 join(6007..6408,1..831)  
 /note="11 (x5 replication, nicking)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03303.1"  
 /db\_xref="GI:215395"  
 /translation="MIDMLVLRPFDLSVGRSLSGNDLAFVDSKIATSGNLSA  
 RVEVHIDCLTVSGLSHPFESLPFTYSGIARFVSGSKNFPCEVETKASPAKVLQGH  
 NVGRTDIALGSEALLNPANSPLCTVLDNATIRIRATFAPAPNEMALACVI  
 DILNWSNQRTSTRSQRMESVTFNENSRHRTIYAKHBLQDQIQLSKSPKAKM  
 TSYKEQLVLSNPDLFASJVFKAIEFVLSKFGPLNFDARFASDVNRQS  
 KOLIFLWSFGFSELFKAFESGNIVDQSAVLQSKHPTPSGCTSFKAARVF  
 GYFRRLVWGSDVVALTPWRFNSFMYVALVECGIPKQQLNLSYCNVPLVPRIV  
 DFSQRPDWVNEPVLKIA"  
 496..831  
 /note="X (function unknown)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03304.1"  
 /db\_xref="GI:215396"  
 /translation="KNYEDSAVLEAPVQSKPTTPSGKTSFAKARVFGFRRVNE  
 GDSVALTPRNSFWRYVALVECGIPKQQLNLSYCNVPLVPRIVNDFSSQRPDW  
 VNEPVLKIA"  
 843..1106  
 /note="V (single-stranded binding protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03305.1"  
 /db\_xref="GI:215397"  
 /translation="KKEIKPSQAGTTRGCVSRQCKPVSLNEDLCIVDLNRYVVL  
 VRTLDEQFAYAPGLYTVHSSFKVGGFGSLMIDRLVPAK"  
 1108..1209  
 /note="VII (morphogenesis)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03306.1"  
 /db\_xref="GI:215398"  
 /translation="WEQVACPDITVCAMIQISVLCFALINAGQR"  
 1206..1304  
 /note="IX (minor coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03307.1"  
 /db\_xref="GI:215399"  
 /translation="SVLVYFASFEVLCMLRSGTTFTRIMETSS"  
 <1301..1551  
 /note="mRNA (3' end +/- 14 bp)"  
 1301..1522  
 /note="VIII (major coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03308.1"  
 /db\_xref="GI:215400"

## CDS

## RNA

## CDS

```

CDS
      1579..2853
            /translat="MKKSLIVKASVAVATLVFMLFSAEGEDPAKKAFFSIOASATVEY
            IGAMVAVVIVGATIGIKLFFKPKSKAS"
            /note="111 (adsorption protein)"
            /codon_start=1
            /transl_table=11
            /protein_id="AA032309.1"
            /db_xref="GI:25540."
            /translation="MKKLLPAPLVVPEVSHSAETVESCLAKRHNSHTVWVLDKPTK
            LDPKAVYEGSLNMAIVGVCTGDEOCQGVTPVGLA.PENEGGSGEGSSEGGSESG
            GGRKPEVIGTVPFGTIVINP.LDGYTPRGTDINPANNPS.LEBSQPLINTPKQNNRPK
            NRQKALVYVGVTVQTDPAVKTYOYVPLSSKAMVAYAKMGKRPDCAFPISGKLEPPFPA
            CEYQGGSSDIPGPVNAAGSGSGSGSGSGSEGGSEGGSGSGSGSGSGSGSGSGSGSGSG
            DYKRMANNAGNTENACENALSDAKGCLSVATYGAALIDGF.CYVSGGLANESGKAT
            GDAFGNSGMAOVAGGCGNSPLMNKPRVY.PSLPOSVECRPYVFGAKPKPEFS.CDCKK
            NLRPGVPAFLVVA.TFMVYFSTFANLNKES"
            2856..3394

```

CDS

```

misc_feature
CDS
        /note=" : (morphogenesis)"
        /codon_start=1
        /transl_table=1
        /protein_id="AAA32311.1"
        /db_xref="GI:215403"
        /translacion="MAVVFVTRKLGSKTIVSGKIQDRVAGCKIAIYLDLRFLOF
        QVRFKNTPLVRLPKRPSIDLAISGNGSYFENKGLVLEEGTWNPSMNDK
        EROELIYSHARKCWMQILVQGLSTVQDASAAAEHVYCRLEITLFGSTL
        YSLVTSKCKPLRLKRGVYKVGDSGLSTFVRRRLVTKLAINVYDRAKAFSSVNDGK
        YSLTLPESHGRFERNLQCKMLTKYLKRFKRVLCALINASKAFSTVYTPKRA
        VKWVQVQIDFQKFTTDSQRNLISYRVFKSGKLNSDLQXGQ611YIDLCT
        S:KKGNSNEVVCNN"
        3817..4081
        /note="internal signal sequence for gene : product 56"
        4221..5501

```

```

misc_feature      3817..4081
                  /note="internal signal sequence for gene : product :61"
CDS
                  4221..5501
                  /note="TV (morphogenesis)"
                  /cdon_start=1
                  /transl_table=1
                  /protein_id="AA03332.1"
                  /db_xref="G:215404"
                  /translation="MKLLNVLNPFVFLNFWSSSEFAQVLENNNSPLRDEPTWYSKOTGE
SVIASEDVAGVTVVSSLDVKKPENLRNFFSVLEANNPDWVGS:PSIIQVNFNSQDY
DELPESDIOEYVNDENAPGSGFPVQNDNVQTKINNVPAKDLRVLELVFNSTKSG
SNVLSVADGCVNLIVGAPDLIDNLPQPLSTGLPTQD:LIEG:IFVQGGALDPSFG
AASQKGTAVGCVNTRDLTSLVSSAGGSFGIENDEY:GLEGRALKNSKGL:LSVPRII
TSGQGG:SVQGVQNPFLITGVTSGLVSSANVNNPCTGERQNVG:SNMVFVAAGNIVV
LITTSKASLSSSTQASDVITNORSITATNN:PDGQTLGLGLITTKYKSSQSGVFP
SKPIPLGILFSRSRNSNESTLVLVAKITVRDL"
rep_origin
                  1578 a 1295 g 22.0 t
                  /note="origin of viral strand synthesis"
Origin            The unique HindIII site.

```

[illegible]

Db 96: TATCCGATCTGTTCAAAATTAACCTCTGAGAAAGGATGACCCAGCATGCGCTGGTCTG 1022  
Qy 161 TACACCGTGCACGACCTGCAAGAGCTGGCGCATTTGGCGAGCGTAAATGACACGGC 240  
Db 1023 TACACCGTGCATCTGCTCGTTCAAAAGTGGTCAGCTCGGCTTCTTTATGATTAACCGT 1082  
Qy 241 CTGCGCTCGTGGTCCCGCGCCAACTAA 264  
Eb 1093 CTGCGCTCGTTCGCGTAACTAA 1106

RESULTS	
AXJ060317	AXJ06017
LOCUS	Sequence 16 from Patent WO9306587.
DEFINITION	DNA
ACCESSION	AXJ06017
VERSION	AXJ06017.1
KEYWORDS	G1:9528946
SOURCE	
ORGANISM	synthetic construct
	synthetic construct
	artificial sequences.
REFERENCE	1
AUTHORS	Ge, L., and Ilaq, V.
TITLE	Novel method and stage for the identification of nucleic acid
JOURNAL	sequences encoding members of a multicentric (poly)peptide complex
	Patent: WO 9306587-A19-1-FEB-1993
	CE LINKING (DB): ILAQ VIC (DB)
FEATURES	LOCATION/Qualifiers
1 source	1..7055

55

CD5

33

CD3

CCS

```

"organism="Synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:12610"
<!-- phage vector from R collection -->
<!-- 831
/feature="C-terminus gene 1"
/codon_start=-
/trans1_table=11
/protein_id="CAC05195.1"
/db_xref="gi:15928947"
/translation="MATTIRINFTSPAPRNENIAKQVIFLWVSGTQTSRSGQ
MESTYIMETSRRETTLVAYIKHVELDHQIQOGLSSKPSAMSTYSQAEQIKVSNPLEL
FASGIVREAEITERRKISGFLPKLFLAIRFASSTYKSGKILFLWFSFSEFLPKR
FEGDSKMINVEDSALVDK10SKKPTIFPSKTSFPAKSRYPGFRDVEGSDVALTY
PNSFWRVSLAVECGIPKSQLMNNLSTCNWVQVYFIVNDSSORPNWYNAPVLIKFL"
496.. 831
/feature="gene X"
/codon_start=-
/trans1_table=11
/protein_id="CAC05156.1"
/db_xref="gi:15928948"
/translation="MNLYDQSLALDLSKSKPTTIFPGQTSFPAKSRYPGFRVLVNE
YDSEALTLMPNPSWFRVSLAVECGIPKSQLMNLSTCNWVPLVYFIVNDSSORPDM
YNEPLTKIA"
843.. 1106
/feature="gene Y"
/codon_start=1
/trans1_table=11
/protein_id="CAC05197.1"
/db_xref="gi:19928949"
/translation="MIVELKPSQCFITPSGYSQKFFYSMLNEQLCYDZGANEYPVIL
VKTTDECGAPVAFGLVYTLFSFKVQGGSLMDRLRLVPK"
1108.. 1203
/feature="gene V1"
/codon_start=1
/trans1_table=11
/protein_id="CAC05198.1"
/db_xref="gi:19928950"
/translation="MECVAFDTITVQAMIQISVLCFALGTLAQGR"
1205.. 1336
/feature="gene X"
/codon_start=-
/trans1_table=11

```

```

CDS
    /protein_id="CAC05199.1"
    /db_xref="GI:9928951"
    /translation="MSVLYVSPASVFLQCLRSQITTFRLMEVSSCVSL"
    1301..1522
    /note="gene V111"
    /codon_start=1
    /transl_table=11
    /protein_id="CAC05200.1"
    /db_xref="GI:9928952"
    /translation="KRSLVTLKASVAVATVYMLSTFAEGDDPAKAPDSLCASATKX"
    1643..2302
    /note="cat resistance gene"
    /codon_start=1
    /transl_table=11
    /protein_id="CAC05201.1"
    /db_xref="GI:9928953"
    /translation="KEKITGYTVTDLSQMHREPERAPQSVACQTYNQQLDITAF"
    LKTVKQKIKFFVPAFHLIARLMAHPEFEMAKDGLVIMDSVPCVTEFDEQETFF"
    ASSLMEHSDPROFLHIYSODVACGEMLAFFKGFLENMFVSNPMWSTSELVAV"
    ANMDFEAFVPEFMKXYTQGDVCMPLAIQVHANVCDGFPHGMNMLEQQICDEMGG"
    A"
    2494..2498
    2517..2521
    2522..2556
    /note="lac operator"
    2607..3407
    /note="ompA FLAG-jun-gene ::lc"
    /codon_start=1
    /transl_table=11
    /protein_id="CAC05202.1"
    /db_xref="GI:9928954"
    /translation="MKCTAIAVALAGFAVAQADYKQVDAGRIARLIEEVYTLXAK"
    QNSLESTAMVLEQVVALQKQMHVGAEEAGGSGGSGSGSGSGSGSGSGSGSGSG"
    FSGGGSGSGSGSGSGGDEYEXKANANKGEMNDENLQSDAKGLDSVATDYGALIT"
    FGAGIDSPAGNAGATGPPAGNSQMGQNSPLMNFROYLBSLPQSVCECPV"
    PSAGKPEPSIDCKXNFRGVFAFLVATFVYVSTFANLIRNES"
    2607..2659
    /note="ompA signal sequence"
    2670..2681
    /note="FLAG peptide"
    2697..2816
    /note="jun peptide"
    2832..3404
    /note="gene III C-terminal domain"
    3446..3488
    /note="f4 terminator"
    3503..3841
    /note="gene VI"
    /codon_start=1
    /transl_table=11
    /protein_id="CAC05203.1"
    /db_xref="GI:9928955"
    /translation="MPVLGGLPLRFLGFLVTFQGYLLFLKGGFGXIALSLF"
    ALIIGNSILVGLSDISKQPSDFQGVCLILPSNMLPFTVYLSVKALIFFDVQX"
    KISYLDWDX"
    3844..4390
    /note="gene I"
    /codon_start=1
    /transl_table=11
    /protein_id="CAC05204.1"
    /db_xref="GI:9928956"
    /translation="MAVYVTVGKSGSGKTVSNGCKDRIVAGCKIATNLDLRCONLP"
    QVRFKPTKPVARIPEKPSISDLAIGGNDSYDEKNGSLVIDEGGTFPNTREKNDK"
    BROPIIDMFLHAKRGMDIFLVQDLSTVDAQASALAEHVYCRRLDILTFVGTIT"
    YSLTLPILSHCRFFKPLNDGQMKLTXYLKKPSRVLCALIGFASAFVSYITCKKE"
    VKKWSYATYPDFDITLDSQRLMSRYVFFDSKGLNSDDQKQGSYITVILCTVKN"
    SIKKQNSNEVKN"
    4868..6148
    /note="gene IV"
    /codon_start=1
    /transl_table=11

```

	protein_id="NC_050525.1"	/protein
	/db_xref="gi:992895"	/db_xref
	/translation="MKHLINVFVDFYFSSSSFRQVIEMKSRFLRFDVNVYSQTSE SVLSPVKQITVYSDVAKEPRLRFIVLRANKPQWDS;PSIIIOXYPNQDYI DESSSIPCEXDNSARSCGFEPQNDNTOTFKINRPADLLRVEI.FKSNITSKS SNVSYGQNILVYSAPKDIQLWJGFSLTFLPDQCLLGILEVCGGLALPESPA AGSGRTAGNUNVNRATSVLSAGGSPETFMGDVALISVALKTNSKSILSPRL TLSCRSKTSYGONPFITERTGESANINMPQRYECNNVS;SNSVPAXACGNIV LSTPKGLSSSSQQASDVLTNGHSATTNVRATTPDCTLLGGILLTKMTSDSGPRIL SKPLTFDRSRSDSEFTLVAKATIVAL"	
misc_signal	6-146..6224	/note="packaging signal"
rep_origin	5264..6403	/note="ori"
CDS	6654..7395	/note="cd ori"
	/note="TN-terminus gene II"	
	/codon_start=1	
	/transl_table=1	
	/protein_id="NC_050526.1"	
	/db_xref="gi:9928958"	
	/translation="VIDMLRLRPTDISVCGRSLSGNDLIAVELSLKATISGKW;S RVEHIDGDLTVGSHPBSLPPIHSIAIKRTIEGSKRFPCVETASPAKKVLQG NVGTTDLALCEALLNFANS;SCPCLYDJLDV"	
BASE COUNT	1769 A 1433 C 1492 G 2357 T	4 others
ORIGIN		
Query Match:	55.2% Score 145.6 DB 6 Length 7055;	
Best local Similarity	72.3%, Pval No 1.4e-13;	
Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;		
Dy	1 AGATCAAGGTGAGAACCCAGGCCAGCCACCAGTTACACCTGGACACGGGCTTAGC 63	
Dd	843 ATGATTAAGTTGAATAATTAAACGCTGCACAGCCCAATTACACACCGCTTGATTTCT 302	
Dy	61 CGCGAGGGGAAGCCCTAAGCCTGAACAGACACTGTGCTACGTGACTGGCGACAG 120	
Dd	903 CGTCAGGGGAGACCTTATTACCTAAAGAACACCTTGATACGTTGATTAGCATTA 962	
Dy	121 TAACCCGTGCTGTGTAAGATCACCTCGAGAGAGGCGACGCCGCTACGCGCGGCTG 180	
Dd	963 TATCGGAGCTCTGTGAAGATTAACCTCGAGAGAGGTACAGCAGCGATACCGCTG 1022	
Dy	161 TACACCGTGCACTGAGACCTTAAAGTGGCCATTTGGCAGCCTCATGTATCACC 240	
Dd	1023 TACACCGTGCATCTCTCTGTTAAAGTTGATCAGTTCCGCTTCTTATGATTACCG 1082	
Dy	241 CTACGCGCTGCTCCGCGCAAGTAA 264	
Dd	1083 CTCGCGCTGTTCCGCGCAAGTAA 1106	
RESULT 6	BD131874	TCS LR DNA Linear PAT 18-SEP-2002
LOCUS	BD131874	
DEFINITION	Novel method and phage for identifying nucleic acid sequence encoding polypeptide complex member.	
ACCESSION	BD131874	
VERSION	BD131874.1 GI:232268.9	
KEYWORDS	CP 2002501721-A/6.	
SOURCE	synthetic construct	
ORGANISM	artificial sequences.	
REFERENCE	1 (bases 1 to 7055).	
AUTHORS	Rudek,F., Ge,R. and Traag,V.	
TITLE	Novel method and phage for identifying nucleic acid sequence encoding polypeptide complex member	
JOURNAL	Patent: JP 2002501721-A 6 22-JAN-2002;	
	MORPHOSYS AG	
COMMENT	OS Artificial Sequence	
	PN CP 2002501721-A/6	
	FD 22-JAN-2002	
	PF GS-AUG-1998 CP 2000505347	
	PR GS-AUG-1997 EP 97113319-4	





```
CDS
496..831
/codon_start=1
/trans_table=11
/product="replication protein px"
/protein_id="AM73898.1"
/db_xref="GI:21667081"
/trans_table="MNIYDSDAVLDAIQSKHFTIPDSKTSAPKASRYFGFRRVNE
GYSVLATMPRNSFWRYVSAVVECGIPKQGLNLSTGNVPLVAFINVDSSQRFCH
INPPTVKIA"
843..1106
/codon_start=1
/trans_table=11
/product="cyclosolic single-stranded DNA binding protein
pv"
/protein_id="AM73897.1"
/db_xref="GI:21667082"
/trans_table="N:KVEIKPSQAQFTIRSVSRQKPEVINEQCCVYDQENYFVL
VKITLDEGPAVAPGLYVHSSFKVQCFQS.MIDRLRLVPAK"
1108..1209
/codon_start=1
/trans_table=11
/product="minor coat protein pxi"
/protein_id="AM73899.1"
/db_xref="GI:21667084"
/trans_table="MSVLVVSFASVFLGMCLSGCTFTRLMETSS"
1301..1522
/codon_start=1
/trans_table=11
/product="major coat protein pviii precursor"
/protein_id="AM73900.1"
/db_xref="GI:21667085"
/trans_table="MKKSLVAKASVAVATLVPLSPAEEDPAAVAATDS-QASATEY
IGYAMAVVIVIGATIGIKLKKF7SKAS"
1579..2853
/codon_start=1
/trans_table=11
/product="adoption protein piii precursor"
/protein_id="AM73901.1"
/db_xref="GI:21667086"
/trans_table="MKKLFAPLPLVPFYSASAEVESCIAKSHTESTFNWKKDKT
LDRYANEGCLMATGVVCTGDETCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
NROGCTLVYGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
CYEKANAKAGAAENADENALQSLAKGKLDVAATVGAATGGF.GVVSGLAMGAA7
GDFAGNSQMAQVQDGNPLMNNFRQYVPSLPSQVECRFPVFCAGKPYEFSDCCX:
NLFRGVAFPLLVATEWVVFSTANILRNKES"
2856..3194
/codon_start=1
/trans_table=11
/product="minor coat protein pvi"
/protein_id="AM73902.1"
/db_xref="GI:21667087"
/trans_table="MPELGIPLLRFLGLLTVTFGYLTLFKKGFCAIAISLF
ALIGNSILVGYLDSIAQLPSDFVQVQLILPSNALPCFVILSVKAAITFVQK
KIVSLDMWK"
3197..4243
/codon_start=1
/trans_table=11
/product="morphogenetic protein pi"
/protein_id="AM73903.1"
/db_xref="GI:21667088"
/trans_table="MAVYEVYGLSGKTVYSGKIDKIVAGCKATYLDLRQNL
QVGRFACTPRYLRIPKPSISDLIA.GRQNDSDZKNGKGLVLDGCTFNTSRKNDK
ERQPTDIFLHARLGDIIFLVQDLSIVDKCARSLAAHVAYCRRLDTLPLFVGL
YSLVTGSKMPLPKLHVGVVYQDLSPTVERWLVTGKXNLVNAYPDKQAFSSNDGIV
```

```
CDS
VSGLTPYLSHGRYKPEKLNQKRLTKYIKRSNVLCTAIGFASAFVSYITOPKPE
VAKVSTQYIDFKFTIDSSQRLNLSYRIVFQDSKGLIKSDLDKQKQSTYIDLCY
SIKQNSSEIVKCN"
4221..5501
/codon_start=1
/trans_table=11
/product="morphogenetic protein lv"
/protein_id="AM73904.1"
/db_xref="GI:21667089"
/trans_table="VKLNVYVFEVDMVSSSSKQVLEMRNSFLKPVYVSKCTGS
SVVSPDVKGVTVVSSQVPEVNNFISVLRANRDMVSGEISIIQKYNRSCDYI
DELPSSDIQEDENASQSGFVPCNVYVOTFTKINVRACILRVLELFVNSKTS
SNVLSVDSGNSLVASPKDILDMPCFSLVLELPRDQITLLELFEVQSDLPSTFA
ASQSGTVAGGVNTRTSVSSAGSSGSENGCIGSVRAKYNRSKLSVPRIL
TISQKGISISQGVNPFITGRVTSANVNPFCVQVQVQVQVQVQVQVQVQVQVQV
LDTISADSSSTVQASVITNQRSLAIVVLRDQCTLLGLTYVNTSCQSGVPEL
SKPLGLFSSRSISNESTLYLVKATIVRAL"
5878..6537
/codon_start=1
/trans_table=11
/product="chaperonin co. acetyl transferase"
/protein_id="AM73894.1"
/db_xref="GI:21667079"
/trans_table="MEKRTGYTVVJISQMRKHEFAFGVQACTYVQVQLDITAF
KTVKKNKREYPAVHILARLNARREFRMAKQGEVITNSVPCVTFHAECEIT
SLMSEYEDDFRQPLHVSQVACGENAVFPKGFENVFVSANPWSFSDLV
AMDNDFAPVFTMGKYTGQKVLMLPLAIVHAFVCTGFHGMMLNLCQVCEBQGG
A"
BASE COUNT 1962 a 1585 c 1641 g 2587 t
ORIGIN
Query Match 55.2%; Score 145.6; DB:2; Length 7775;
Best Local Similarity 72.0%; Pred. No. 1.3e-13;
Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
CDS
1 ATGATCAAGTGAAGTCAAGCCGACGCCGCCGCTTACCAACCCGACGCCGCTGAGC 60
2 843 ATGATTAAGTTGAAATTAACCGTTCAGGCGCAATTAACCGCTTCTGTTCT 902
3 61 CGCAGGCGGACGCTACGCTTAAGCGAGCGCTGTCTAGCTGAGCTGCGACGAG 120
4 903 CGTCAGGCGACGCTTATCTAGTGAATGACGCTTTTCTAGCTGATGATGAA 962
5 121 TACCCGCTGCTGTAAGATCAACCTGTCAGAGGCGCGACCCGCTAGCCCGGCTG 180
6 363 TATCCGCTGCTTGTCAAGATTAATCTGCGAGGACGACCGCTATGCGCTGCTG 1022
7 181 TACACCGTCACTGACGCGCTCAAGGTCGCGCGCTGCGACGCTATGATGACGCG 240
8 1023 TACACGCTGATGTCGCTGCTGCTCAAGGTCGCTGCTGCTGCTGCTGCTGCTG 1082
9 241 CTGCGCTGCTGCTGCGCGCGCAATGA 284
10 1083 CTGCGCTGCTGCTGCGCTGAATGA 1106
RESULT: 6
AX006302 7793 bp DNA linear PAT 24-NOV-2000
LOCUS AX006302
DEFINITION Sequence 3 from Patent WO9906587.
ACCESSION AX006302
VERSION AX006302.1 GI:1992893;
KEYWORDS
SOURCE
ORGANISM
REFERENCE
Ge. and Jia, V.
AUTHORS Novel method and phase for the identification of nucleic acid
CITILE sequences encoding members of a multimeric (poly)peptide complex
JOURNAL Patent: WO 9906587-A 3 11-FEB-1999;
FEATURES
location/Qualifiers
```

```
source
1..7793
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/feature="phage vector (circular)"
<1..831
/feature="C-terminus gene 11"
/codon_start=1
/translation="NATTSRIDATSPARAPNENAKGVICHURVNSNGOTSTSSCN
WESTVWNSRRTVVAHLKHELQO:QCSKSKAKTSYQKQLVLSNPDLLE
PAGLVPEARLETRYLKSFGJLNLDAIRFSDJNRKQKDLFDLWFSFSEJKA
FEGSMMIYDGSAYDAIQSXHPITPSGKTSFSAKSRVFGFRRJLVNBSYDVALTM
PNSFMRYVSALVEGCI:PKSQLMNLSCNNVPLVIFINVDPSQRFDWYNBPVLKIA
"
496..831
/feature="gene X"
/codon_start=1
/translation="CACOS184.1"
/protein_id="CACOS184.1"
/db_xref="GI:9928933"
/translation="MNIYDGSALVDALQSKHFTTFSKTSFAKASRYGFRRLVNE
GYDVALTMPNSFMRYVSALVEGCI:PKSQLMNLSCNNVPLVIFINVDPSQRFDW
YNBPVLKIA"
843..1106
/feature="gene V"
/codon_start=1
/translation="CACOS185.1"
/protein_id="CACOS185.1"
/db_xref="GI:9928934"
/translation="MIKVEIKPSQAOFTTSRGVRCGKPYSLNEQLCYVDLNEYVYL
VKITDEGPPAYALGLTYHLSSFKVGFQSLMIDRLRLVPAR"
1198..1209
/feature="gene VII"
/codon_start=1
/translation="CACOS186.1"
/protein_id="CACOS186.1"
/db_xref="GI:9928935"
/translation="MEQVADFDITVQMIQ:SVWLGFALG:INGGQR"
1206..1316
/feature="gene IX"
/codon_start=1
/translation="MSVLVVSFASFV:GMCLRSGLTYFTRLMETSSCVSL"
1301..1522
/feature="gene VII"
/codon_start=1
/translation="CACOS188.1"
/protein_id="CACOS188.1"
/db_xref="GI:9928937"
/translation="VRSKLVLKASVAVAT:VPMLSFAAGSDPAKKAAPSLQASATRY
IGVAMWVIVGATIGIKLFKRTSKAS"
1643..2302
/feature="cat resistance gene"
/codon_start=1
/translation="CACOS189.1"
/protein_id="CACOS189.1"
/db_xref="GI:9928938"
/translation="MEKKITCYTVQD:SCMRKHEFEAFQSVACQ:VYNOTVQID:CAF
ZKTVKAKHKFPYPAFHILARLMAHFEFFMAKDDGLVWDSVHPTVFHQTETP
SLSWSEHDDRFOP:HLIYSDQVACVGNLAYFPKGFLENMFVVSANPWSFTSFDLNV
ANMNDPFAVPTMGKYVTOGDKVLPLAIOVHAHVCDFGVHGMNLELOYCDMEWCGG
A"
2656..2669
/feature="10'-signal"
2679..2683
/feature="1ac operator"
2769..4139
/feature="ompA-FLAG-scfv (anti-HAG)-gene 11ss"
2769..4139
```

```
/codon_start=1
/translation="CACOS190.1"
/protein_id="CACOS190.1"
/db_xref="GI:9928939"
/translation="VKKRTATLAVLAGFAVACADYKCTWTOSSRSLLTATGKWT
VSCITSSQCSLNSGKQKNTWQKFGDPKVLITWAS:RESGVDFRTGSSGCTDFT
LITTSVQAEDELAVY:CONVSNPCT:FGNSTKLELFAAAGGSG:KSGGSGGSGG
GGSGGGGSEGLVYESGGD:LVNPGSG:KLSCASGSEFSYGMVWVCTPPDXRLEWVA
TISNGGYTYVPSGVKGRFTISPNKAKSLYLQMSGLKSGDSAYVYCARERVDXGA
AWGCGGLVTVSASGEFSAAGGSGSGGDFDEKXANAKKAVETNADKALQSDA
KGLISVATVGSAA:IDGFT:GDVSGLWNGAG:GTFRGSNSQVACVGGGNSLMMNPR
CXT:PSL:POSVECFPPVFGAKKPIRESIDCKRINLRFQVAFILLYATRWYVSTFANK
LRKRES"
2769..2831
/feature="ompA signal sequence"
2832..2843
/feature="FLAG peptide"
2844..3641
/feature="scfv (anti-HAG)"
3666..4139
/feature="gene 11'ss"
4178..4220
/feature="fd terminator"
4231..4569
/feature="gene V"
/codon_start=1
/translation="CACOS191.1"
/protein_id="CACOS191.1"
/db_xref="GI:9928940"
/translation="MPVLGIPILLFJGLVLT:FGVLLTF:KSGFGKIAIA:SLFL
ALIG:NSILVGLSD:ISADPEEFVQGVQLILPSVALPQFVILSVKAAIFFDVQK
KIVSLDMDK"
4572..5638
/feature="gene I"
/codon_start=1
/translation="CACOS192.1"
/protein_id="CACOS192.1"
/db_xref="GI:9928941"
/translation="VAVYFVGTGLSGKTVSGK:GDYVAGCKIATYLDRL:QNTLP
OVGRFAKTPPV:RIPDPSISDLA:GRGNDSYTEFKK:LVN:DEGCTFMNRSKMK
BROPIIDRFLHAKKLMDI:FLYQD:SLYDKOKRSLAHVYGRLEDTLP:RYVETL
YSKTSKMP:PKL:VGVKRTGDS:SLSPVVERLVYK:LVNAYD:KAK:SSRISBVG
YSILTPYLSHRYFKPLNIGQMK:KIVAKRSRYLCA:IGFAS:FTY:SYITQPPS
VKVYSCITVPEKFTIIDSORLNL:SYRYVFKSGKGLINSDELQKQSVITV:DLCTV
S:KXKGNSEIYKCN"
5558..6876
/feature="gene IV"
/codon_start=1
/translation="CACOS193.1"
/protein_id="CACOS193.1"
/db_xref="GI:9928942"
/translation="VKLVNINPV:LVFVSSSEFPAQV:EVNNSP:RHFPTWVSQKQGS
SVIYSPVKQTV:VYSSDVKPEVLRNFISVLKANEVWGSIP:CYVNSQCYI
DELPSR:CEYDNSAPSGGFV:PDNDVOTFKIKVNAKQ:RYDEL:PVNSQYI
SMIYVNGSL:LVSAKCT:LMN:PSFLSNOLPTKOC:IL:EG:EFVCGDAL:DEFA
AGSBRGVACGVYTDRLTSL:LSAGSGF:FNQVY:GLSVRAK:VNSAKXIT:SVRPI
TLSQKGISVGGVNP:ITGRV:GESANVNPQ:VERQNV:SMVFPV:VNAAGCNIV
LITISKADS:SSSTOASDVITNORSIA:TVNLDDGTL:LLGGITLVKXNS:QDSGVPT
SKIP:IGL:FSRSSDSNEESTLVYLVXATIVRAJ"
6874..6952
/feature="packaging signal"
6992..7117
/feature="fd ori"
7382..7793
/feature="N-terminus gene 11"
/codon_start=1
/translation="CACOS194.1"
/protein_id="CACOS194.1"
/db_xref="GI:9928943"
/translation="VIDMLVRL:PLFIDS:LVCSR:LSGNDLIA:VYDLK:ITLSGAMISA
RVEYHIDGLTVSGLSHPPE:LP:HYSG:AFK:YEGSKNF:PCVE:IKASPAKVLOGH
NVFOITDLACSEALLINPANS:PLC:YDLDV"
BASE COUNT 1920 a 1649 c 1667 g 2543 t 4 others
```

## ORIGIN

Query Match 55.2%; Score 145.6; DB 6; Length 7783;  
 Best Local Similarity 72.0%; Pred. No. 1.3e-13;  
 Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCCAATTCACACCCGAGCGGCTGAGC 60  
 DB 843 ATGATTAAGTTAAATTAACCGTCTCAAGCGCAATTAACACCGCTTGTGTCT 902  
 QY 61 CGCCAGGCGAAGCCCTACAGCCCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 DB 903 CGTCAGGCGAAGCCCTATTCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
 QY 121 TACCCGCTGTGTGAGATCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 DB 963 TATCCGCTGTGTGAGATCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022  
 QY 181 TACACCTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 DB 1023 TACACCTGATGT 1082  
 QY 241 CTGGGCTGT 264  
 DB 1083 CTGGGCTGT 1106

## RESULT 11

SD131871

## LOCUS

BD131871

## DEFINITION

BD131871

## ACCESSION

BD131871

## VERSION

BD131871.1

## KEYWORDS

JP 2002501721-A/3

## SOURCE

JP 2002501721-A/3

## ORGANISM

JP 2002501721-A/3

## REFERENCE

JP 2002501721-A/3

## AUTHORS

JP 2002501721-A/3

## TITLE

JP 2002501721-A/3

## JOURNAL

JP 2002501721-A/3

## COMMENT

JP 2002501721-A/3

## OS

JP 2002501721-A/3

## PN

JP 2002501721-A/3

## PD

JP 2002501721-A/3

## PE

JP 2002501721-A/3

## PR

JP 2002501721-A/3

## PI

JP 2002501721-A/3

## PC

JP 2002501721-A/3

## CC

JP 2002501721-A/3

Novel method and phase for identifying nucleic acid sequence  
 encoding polypeptide complex member.  
 BD131871  
 BD131871.1 GI:23226816  
 JP 2002501721-A/3  
 SYNTHETIC CONSTRUCT  
 SYNTHETIC CONSTRUCT  
 ARTIFICIAL SEQUENCE  
 1 (bases 1 to 7783)  
 Rudert, F., Ge, R. and Irig, V.  
 Novel method and phase for identifying nucleic acid sequence  
 encoding polypeptide complex member  
 Patent: JP 2002501721-A 3 22-JAN-2002;  
 MORPHOSYS AG  
 OS Artificial Sequence  
 PN JP 2002501721-A/3  
 PD 22-JAN-2002  
 PE 03-AUG-1998 JP 2000505327  
 PR 01-AUG-1997 JP 97113319.4  
 PI PRITZ ROBERT, RIMIN GE, VICK IRIG  
 PC C12N15/09, C12N15/00  
 CC Description of Artificial Sequence: phase vector phagA CC  
 (circular)  
 CC C-terminus gene 11  
 CC gene X  
 CC gene Y  
 CC gene VI  
 CC gene VII  
 CC gene VIII  
 CC gene IX  
 CC cat resistance gene  
 CC ompA-FLAG-scfv (anti-HAG)-gene 11ss  
 CC FLAG peptide  
 CC scfv (anti-HAG)  
 CC gene 11ss  
 CC gene VI  
 CC gene VII  
 CC gene VIII  
 CC N-terminus gene 11  
 CC lac operator  
 CC fd ori  
 CC packaging signal  
 CC fd terminator

## FEATURES

FT Key Location/Qualifiers  
 FT CDS 1.. (828)  
 FT CDS 14961.. (1828)  
 FT CDS (843).. (11103)  
 FT CDS (11081).. (12061)  
 FT CDS (11261).. (1313)  
 FT CDS (11301).. (1519)  
 FT CDS (11321).. (1299)  
 FT CDS (12643).. (4336)  
 FT CDS (2769).. (2812)  
 FT misc\_feature (2812).. (2843)  
 FT misc\_feature (2843).. (2844)  
 FT misc\_feature (2844).. (3641)  
 FT misc\_feature (3641).. (4139)  
 FT CDS (4231).. (4366)  
 FT CDS (4572).. (4645)  
 FT CDS (4596).. (4673)  
 FT CDS (7382).. (7793)  
 FT -10\_signal (2679).. (2693)  
 FT -35\_signal (2693).. (2693)  
 FT misc\_feature (2693).. (2693)  
 FT rep\_origin (2693).. (2718)  
 FT misc\_signal (2718).. (2718)  
 FT terminator (4178).. (4220)  
 Location/Qualifiers  
 .. 7783  
 /organism="synthetic construct"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:12630"

## BASE COUNT

ORIGIN

Query Match 55.2%; Score 145.6; DB 6; Length 7783;  
 Best Local Similarity 72.0%; Pred. No. 1.3e-13;  
 Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCCAATTCACACCCGAGCGGCTGAGC 60  
 DB 843 ATGATTAAGTTAAATTAACCGTCTCAAGCGCAATTAACACCGCTTGTGTCT 902  
 QY 61 CGCCAGGCGAAGCCCTACAGCCCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 DB 903 CGTCAGGCGAAGCCCTATTCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
 QY 121 TACCCGCTGTGTGAGATCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 DB 963 TATCCGCTGTGTGAGATCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022  
 QY 181 TACACCTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 DB 1023 TACACCTGATGT 1082  
 QY 241 CTGGGCTGT 264  
 DB 1083 CTGGGCTGT 1106

## RESULT 12

AF362081

## DEFINITION

AF362081

## ACCESSION

AF362081

## VERSION

AF362081.1

## KEYWORDS

GI:14485027

## SOURCE

AF362081

## ORGANISM

AF362081

## REFERENCE

AF362081

## AUTHORS

AF362081

## TITLE

Reptide display vector (anti-HAG)-gene 11ss  
 Reptide display vector (anti-HAG)-gene 11ss  
 Artificial sequences, vectors.  
 1 (bases 1 to 8233)  
 Bushell-Seifert, D., Stel'vanski, M. and Gershenov, D.M.  
 The rational design of a "type 89" genetically stable peptide  
 display vector in the filamentous bacteriophage phi  
 Nucleic Acids Res 29 (10), 506 (2001)  
 2 (bases 1 to 8233)  
 Bushell-Seifert, D.



AUTHORS Zinder, N.E. and Horvath, K.  
TITLE Multi-regulatory element of filamentous bacteriophages  
JOURNAL Microbiol. Rev. 49 (2), 101-106 (1985)  
MEDLINE 85240391  
PUBMED 289671  
REFERENCE 4 (bases 1 to 9183)  
AUTHORS Smith, G.P.  
TITLE Filamentous phage assembly: morphogenetically defective mutants  
JOURNAL Virology 167 (1), 156-165 (1988)  
MEDLINE 89045639  
PUBMED 3183194  
REFERENCE 5 (bases 1 to 9183)  
AUTHORS Smith, G.P.  
TITLE Filamentous phage vector fd-ec  
JOURNAL Unpublished  
REFERENCE 6 (bases 1 to 9183)  
AUTHORS Smith, G.P.  
TITLE Direct Submission  
JOURNAL Submitted (16-DEC-1993) Division of Biological Sciences, University of Missouri, Tucker Hall, Columbia, MO 65211-7400, USA  
COMMENT Filamentous phage cloning vector constructed by splicing the tetracycline resistance determinant of T10 into the minus-strand origin of bacteriophage fd. Minus strand synthesis is slow, so double-stranded RF copy number is low (about 1 in growing cells; 100 in stationary-phase cells). Yield of physical particles is about 5e11 virions/ml. Transduce tetracycline resistance into infected host cell. Make very small plaques. Foreign inserts can be spliced into unique HindIII site, and are generally stable.

FEATURES  
source  
1..9183  
Location/Qualifiers  
1..9183  
/organism="Filamentous phage cloning vector fd-ec"  
/mol\_type="other DNA"  
/db\_xref="taxon:127516"  
/lab\_host="Escherichia coli"  
/focus  
1..3645  
/organism="Enterobacteriaceae phage fd"  
/mol\_type="other DNA"  
/db\_xref="taxon:13664"  
8421..9183  
/organism="Enterobacteriaceae phage fd"  
/mol\_type="other DNA"  
/db\_xref="taxon:10864"  
Join(8782..9183,1..831)  
/codon\_start=1  
/transl\_table=1  
/product="replication protein pIII"  
/protein\_id="AA75600.1"  
/db\_xref="GI:8489183"  
/translation="MIDMLVLRPFDLSVGRSLNSGNDIIAFVDSKIAATLSGNLSA  
RTVEYHDDGLTVSGSLHPESLPYVSGAFKIVGSGNFYPCVEIKASPAKVIQGH  
NVEGTDILALCEALILNFANSJCLYDLDVAATISRIDATFAPAAVNEINIAQV  
DHLRNVNCGTQKSTRSQNMESTVNTNSTRRLVAVLIGVLELOHIOQLSSPEAKQ  
TSYQKEQLKVLNSPDLLEFASGLVREARLIERLTKLSPNLPDARFADYRQG  
KDIPLMSFSFSELKFAFGDSNIVYDSAVDAIOSKFTTSPSGTSFPAKART  
GFRRLVNEGYSVALTMPRNSFWRVYSALVECGIPKQDLMMSTOMNVPVLRFINV  
DRESSQRDMNEVLEKIA"  
496..831  
/codon\_start=1  
/transl\_table=1  
/product="replication protein pX"  
/protein\_id="AA75601.1"  
/db\_xref="GI:8489183"  
/translation="NNIYDSSAVLDAIQKRPFTTPSGTSPAKASRFGGFRLVNE  
GYDSVALTMPRNSFWRVYSALVECGIPKQDLMMSTOMNVPVLRFINVDSGSRPDM  
YNEPVLKIA"  
843..1106  
/codon\_start=1  
/transl\_table=1  
/product="cytosolic single-stranded DNA binding protein  
pV"  
/protein\_id="AA75603.1"

CDS  
/db\_xref="GI:8489184"  
/translation="MKETVRSQACQFTTRSGVSGQGRFVSENEICVVDQNEYPVL  
VKITLPEQOPAVAPGLYTVLSSFRVQSGSLMIDRLVLPK"  
1188..1209  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pVII"  
/protein\_id="AA75604.1"  
/db\_xref="GI:8489185"  
/translation="VEQVAFDTIVQAMQLISVLCFALGIIAGGR"  
1206..1304  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pX"  
/protein\_id="AA75605.1"  
/db\_xref="GI:8489186"  
/translation="SVLVYSFASFVSGCLRSQITFTRLMETSS"  
1301..1522  
/codon\_start=1  
/transl\_table=1  
/product="major coat protein pVIII precursor"  
/protein\_id="AA75598.1"  
/db\_xref="GI:8489179"  
/translation="MKSLVLSASVAVATVPELSSFAEGGCFARAAPDSLQASATY  
GVAMWVIVQATGKLFKFKTSKAS"  
1301..1369  
/codon\_start=1  
/transl\_table=1  
/product="description protein pIII precursor"  
/protein\_id="AA75609.1"  
/db\_xref="GI:8489190"  
/translation="MKKLFALPLVVPFYSHAEVSGTAKPFTNSFTVWKCKT  
LRYAVNFGCLNNAQVAVVCTGDFCCQGVKPLGLAIPENGSGSGSGSGSGSG  
SGTRPEPGDTPFPGTYINPLDGTTPPTPEPMAPNPSLESGSPPLNTPKQNNRFR  
NRQGLCVITGTVTQSTDPKTYVQTVVSSKAMPDAVANKGFRQCAFESGNEDEPFR  
CEYQGGSDLPQPPVNAAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG  
DYKXNNAKNGAMTENDALQSDAKGLDSVATLVGAALGGTGVVSGALNGCAT  
GPAFGSSQMAQVCGDNGSPFLNNRQVLPSPQGVCEHPVYVQAKFYEPFIDCDK  
NLRQGFALIVATGTFMYVSTFALNRKES"  
1579..1632  
2856..3394  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pVI"  
/protein\_id="AA75606.1"  
/db\_xref="GI:8489187"  
/translation="VWVLLGSLFLLRFGQLVTPGVLFLKKGFKMAIATSLPL  
ALIGNSLIVGLSDISAKLPSPFVQGVCLLPBALCPVILSVKAFIFPVKQ  
KIVSYLWDR"  
3197..4243  
/codon\_start=1  
/transl\_table=1  
/product="morphogenetic protein pI"  
/protein\_id="AA75599.1"  
/db\_xref="GI:8489188"  
/translation="SAVIVVIGKLSGSKTVSVKQCKQIVAGCKIAINLDIRLONLP  
QVGRFAKTPRVLRIPIKPSISDLIAIGRNDSDYDNKQGLVLDSCGTFFNRSWMDK  
ERQPIDWFHARKLGWDILFLVODISIVDQKALAEHVYCPRLRITLPVGLV  
YSLVGSKMLPRCHGVAVVYGDQSLSPVSEMLYTKGLVYVADKQAFNSYDGGV  
YSVLTYLSHGRVFKPNIGQKMKTKILKFSVLGIAISFASAPFTSYVTKRPE  
VKKVQSTQPDPKFTIDSGRANLSRVYVFKSKKSLNSDILQKQSGSTIYIDLCV  
SIRKNSNEIVKCN"  
4221..5501  
/codon\_start=1  
/transl\_table=1  
/product="morphogenetic protein IV"  
/protein\_id="AA75602.1"  
/db\_xref="GI:8489183"  
/translation="KKLVNINPFTSRSSSSPQVEMANSYRPTFTWYSKCTGS  
STVSPDVAGTIVVSSDEVKPNLNFTISVLRANPQVGSIPSIQKVNNSQDIY  
DELPSDIOEDDNSAPSGFVPQDNVCTGFKINNVPKQILRVEELFVNSNKS  
SNVLSDVSGNLIVSAPKQILNLDPQISTVCLPTDQILBGLIPEVOGQDALPSFA

CDS  
/db\_xref="GI:8489184"  
/translation="MKETVRSQACQFTTRSGVSGQGRFVSENEICVVDQNEYPVL  
VKITLPEQOPAVAPGLYTVLSSFRVQSGSLMIDRLVLPK"  
1188..1209  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pVII"  
/protein\_id="AA75604.1"  
/db\_xref="GI:8489185"  
/translation="VEQVAFDTIVQAMQLISVLCFALGIIAGGR"  
1206..1304  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pX"  
/protein\_id="AA75605.1"  
/db\_xref="GI:8489186"  
/translation="SVLVYSFASFVSGCLRSQITFTRLMETSS"  
1301..1522  
/codon\_start=1  
/transl\_table=1  
/product="major coat protein pVIII precursor"  
/protein\_id="AA75598.1"  
/db\_xref="GI:8489179"  
/translation="MKSLVLSASVAVATVPELSSFAEGGCFARAAPDSLQASATY  
GVAMWVIVQATGKLFKFKTSKAS"  
1301..1369  
/codon\_start=1  
/transl\_table=1  
/product="description protein pIII precursor"  
/protein\_id="AA75609.1"  
/db\_xref="GI:8489190"  
/translation="MKKLFALPLVVPFYSHAEVSGTAKPFTNSFTVWKCKT  
LRYAVNFGCLNNAQVAVVCTGDFCCQGVKPLGLAIPENGSGSGSGSGSGSG  
SGTRPEPGDTPFPGTYINPLDGTTPPTPEPMAPNPSLESGSPPLNTPKQNNRFR  
NRQGLCVITGTVTQSTDPKTYVQTVVSSKAMPDAVANKGFRQCAFESGNEDEPFR  
CEYQGGSDLPQPPVNAAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG  
DYKXNNAKNGAMTENDALQSDAKGLDSVATLVGAALGGTGVVSGALNGCAT  
GPAFGSSQMAQVCGDNGSPFLNNRQVLPSPQGVCEHPVYVQAKFYEPFIDCDK  
NLRQGFALIVATGTFMYVSTFALNRKES"  
1579..1632  
2856..3394  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pVI"  
/protein\_id="AA75606.1"  
/db\_xref="GI:8489187"  
/translation="VWVLLGSLFLLRFGQLVTPGVLFLKKGFKMAIATSLPL  
ALIGNSLIVGLSDISAKLPSPFVQGVCLLPBALCPVILSVKAFIFPVKQ  
KIVSYLWDR"  
3197..4243  
/codon\_start=1  
/transl\_table=1  
/product="morphogenetic protein pI"  
/protein\_id="AA75599.1"  
/db\_xref="GI:8489188"  
/translation="SAVIVVIGKLSGSKTVSVKQCKQIVAGCKIAINLDIRLONLP  
QVGRFAKTPRVLRIPIKPSISDLIAIGRNDSDYDNKQGLVLDSCGTFFNRSWMDK  
ERQPIDWFHARKLGWDILFLVODISIVDQKALAEHVYCPRLRITLPVGLV  
YSLVGSKMLPRCHGVAVVYGDQSLSPVSEMLYTKGLVYVADKQAFNSYDGGV  
YSVLTYLSHGRVFKPNIGQKMKTKILKFSVLGIAISFASAPFTSYVTKRPE  
VKKVQSTQPDPKFTIDSGRANLSRVYVFKSKKSLNSDILQKQSGSTIYIDLCV  
SIRKNSNEIVKCN"  
4221..5501  
/codon\_start=1  
/transl\_table=1  
/product="morphogenetic protein IV"  
/protein\_id="AA75602.1"  
/db\_xref="GI:8489183"  
/translation="KKLVNINPFTSRSSSSPQVEMANSYRPTFTWYSKCTGS  
STVSPDVAGTIVVSSDEVKPNLNFTISVLRANPQVGSIPSIQKVNNSQDIY  
DELPSDIOEDDNSAPSGFVPQDNVCTGFKINNVPKQILRVEELFVNSNKS  
SNVLSDVSGNLIVSAPKQILNLDPQISTVCLPTDQILBGLIPEVOGQDALPSFA



Fri Nov 7 15:26:53 2003

us-09-622-500b-3\_1.rge

Page 17

Search completed: November 7, 2003, 12:52:43  
Job time : 1511 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 12:27:20 : Search time 1315 Seconds

(without alignments)  
4879.378 Million cell updates/sec

Title: US-09-622-500B-3

Perfect score: 264

Sequence: 1 atgacaaagtgagatcaaa.....gcccgtgtccgcgaagaa 264

Scoring table: Gapped 10.0, Gapex: 1.0

Searched: 2278.392 seqs, 1215223856 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST.\*  
1: em\_estdb.\*  
2: em\_estnum.\*  
3: em\_estlin.\*  
4: em\_estm.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estnum.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pla.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_prc.\*  
25: em\_gss\_rcd.\*  
26: em\_gss\_pbg.\*  
27: em\_gss\_vrt.\*  
28: gb\_gss.\*  
29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length DB	Description
1	51.2	19.4	756 13 BU449043 603765817
2	50.4	19.1	493 9 AW269906 AW269906 xv37d01.x
3	50.4	19.1	493 9 A1971889 A1971889 ww29902.x
4	50.4	19.1	518 9 A1056863 A1056863 oz03ec01.x

7

5	49.4	18.7	427 10 B011441 7093603.x
6	49.4	18.7	433 14 C0828249
7	49.4	18.7	545 13 B0080051 946149508
8	49.2	18.6	454 10 B6311865 WHE2063.C
9	49.2	18.6	585 10 BE436680 WHE1274.B
10	49.2	18.6	596 12 B2351883
11	49.2	18.6	947 13 B3388416
12	49.2	18.6	945 23 B2566686
13	48.4	18.3	974 13 B0412619
14	48.4	18.2	584 29 B2551647
15	47.8	18.1	407 14 C0833570
16	47.8	18.1	443 14 C0885926
17	47.8	18.1	467 13 B0773275
18	47.8	18.1	468 13 B0493439
19	47.8	18.1	471 13 B0278825
20	47.8	18.1	479 13 B0037392
21	47.8	18.1	483 10 B5757529
22	47.8	18.1	487 13 B0779771
23	47.8	18.1	492 13 B0499994
24	47.8	18.1	493 12 B3398391
25	47.8	18.1	493 13 B0073515
26	47.8	18.1	501 12 B1675202
27	47.8	18.1	504 13 B0615678
28	47.8	18.1	511 13 B0778982
29	47.8	18.1	515 12 B1478947
30	47.8	18.1	517 13 B0779374
31	47.8	18.1	519 13 B0487062
32	47.8	18.1	521 13 B0487067
33	47.8	18.1	522 13 B0079594
34	47.8	18.1	534 14 C0885925
35	47.8	18.1	535 13 B0093574
36	47.8	18.1	539 13 B0093535
37	47.8	18.1	539 13 B0093761
38	47.8	18.1	543 13 B0397824
39	47.8	18.1	546 12 B0248476
40	47.8	18.1	546 13 B0293534
41	47.8	18.1	549 13 B0837923
42	47.8	18.1	549 13 B0092494
43	47.8	18.1	551 14 C0605096
44	47.8	18.1	553 13 B0635815
45	47.8	18.1	554 13 B0779073

## ANALYSES

RESULT 1  
BU449043 756 bp -RNA linear EST 29-NOV-2002  
DEFINITION 603765817F1 CSE08RN13 Gallus gallus CDNA clone CSE08T691b5 5', mRNA

ACCESSION BU449043  
VERSION BU449043.1 GI:25938354  
KEYWORDS EST  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

REFERENCE  
AUTHORS Boarman, P.E., Sanz-Buqueiro, J., Overton, I.M., Butt, D.W., Rosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
TITLE A Comprehensive Collection of Chicken cDNAs  
JOURNAL Curr Biol. 12 (22), 1965-1969 (2002)  
MEDLINE 2335534  
PUBMED 12445392

COMMENT  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology (UMIST)  
PO Box 88, Manchester, M60 1GD, UK  
Tel: 01612008930  
Fax: 01612360409

FEATURES  
Source

Email: Simon.Hubbard@nist.ac.uk.

Location/Qualifiers  
1..756

/organism="Gallus gallus"  
/mol\_type="RNA"  
/strain="layer"  
/db\_xref="taxon:9031"  
/clone="CHEST69:1b6"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH:03"  
/clone\_id="CSECERN1.3"

/note="Organ: ovary; Vector: pBluescript II KS(+); Site:1; EcoRI; Site:2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT  
ORIGIN

142 a 234 c 238 g 142 t

Query Match  
Best Local Similarity 53.5%; Pred. No. 0.4; Length 756;  
Matches 107; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

CY 9 GGTGAGATCAAGCCGACGAGCCGAGTCCACACCGCGCGGTAGAGCCGACGAG 58  
DB 169 GGTGAGAGCAACCCCAACCAACAGAGTGGCGGCTACGGCGTTTATCTGAGCCG 226  
CY 69 CAAGCCCTACAGCCTGACACGACGCTGCTACGTGTAACCTGAGCAACAGTACCCGCT 128  
DB 229 CCAAGCTATCTCTGCTGTAAGCAAGGTTTGTGGAAAGAGGCAACCTGAGTGGT 282  
CY 129 GCTGATGATGATCAAGCTGACGAGCGCGACGCGCTACGCGCGCGCGCTGACAGCT 188  
DB 289 GCAAGCGGCGGCTGAGAGTGAAGATTTTCGGAGCGCTACCCCGCTTCCCGCCATCCA 348  
CY 159 GCAAGCTGAGCAGCTTCAAG 208  
DB 349 GTCCATCCAGCTCTCCAGG 368

RESULT 2  
LOCUS AM269906  
DEFINITION Xv73D0.X1 Soares NFL T\_GRC\_S1 Homo sapiens cDNA clone IMAGE:2815297 3' similar to TR:073884 C73884 PUTATIVE PHOSPHATASE.  
/contains TARI.13 TARI repetitive element.; mRNA sequence.  
ACCESSION AM269906  
VERSION AM269906  
KEYWORDS EST. GI:6656936  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 489)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.  
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL CONTACT: Robert Strausberg, Ph.D.  
COMMENT Email: cga@bbs-rcmail.nih.gov  
This clone is available royalty-free through LNC; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand

FEATURES  
Source

Seq primer: 43UP from Gdbco  
High quality sequence stop: 445.  
Location/Qualifiers  
1..489

/organism="Homo sapiens"  
/mol\_type="RNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2815297"  
/lab\_host="DH:08"  
/clone\_id="CSECERN1.3"

/note="Organ: pooled; Vector: pTZ19-Pac (Pharmacia) with a modified polylinker Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung Nbl13W, testis NHT, and B-cell NCI-CCAP GCE) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297483-302087, 692632-697239, 726408-728711, and 728396-731399. Subtraction by Benito Soares and M. Patricia Bonaldo." 1 others

BASE COUNT  
ORIGIN

77 a 181 c 153 g 77 t

Query Match  
Best Local Similarity 51.3%; Pred. No. 0.55;  
Matches 117; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

CY 34 CAGTTACACACCCGCGAGCGGCTGACCGCCGACGACACCTTACAGCTTACAGCAG 93  
DB 188 CCGTTCCACACACAGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 247  
CY 94 CTGTGATGATGATCAAGCTGACGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 153  
DB 248 AGGACGACTGCTGCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 307  
CY 154 GCGCAGCGCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 213  
DB 308 GCGCAGCG 367  
CY 214 CAGTTGCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 261  
DB 368 CCGCGCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 415

RESULT 3  
LOCUS A1971889  
DEFINITION Xv73D0.X1 NCI CGAP C73884 Homo sapiens cDNA clone IMAGE:2815297 3' similar to TR:073884 C73884 PUTATIVE PHOSPHATASE.; mRNA sequence.  
ACCESSION A1971889  
VERSION A1971889  
KEYWORDS EST. GI:5766715  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 493)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.  
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL CONTACT: Robert Strausberg, Ph.D.  
COMMENT Email: cga@bbs-rcmail.nih.gov  
This clone is available royalty-free through LNC; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand

CONTACT: Robert Strausberg, Ph.D.  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Benito Soares, Ph.D. cDNA Library Arrayed by: Carista Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution: information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bdip/image/image.html



temperature on filter paper with water, fystatin and cefotaxime in covered crystallization dishes. Roots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pluscript phageids in the 3' Close Lab (Choi, Close, Fenlon) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the CD Anderson Lab (all other authors).

BASE COUNT 119 a 209 c 160 g 97 t

Query Match 18.6%; Score 49.2; DB 10; Length 585;

Best Local Similarity 53.0%; Pident No. 1;

Matches 105; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Db 44 CCCGACGCGCGGTGACCGCCAGCGGCAAGCCCTGACAGGAGAGAGCTGTGCTACG 103  
106 CACGATTGAGAGTGGCCCCCAAGCCCAACGCTTATGCGGTGGGAAAGGAGCCACT 165  
Qy 104 TGAACCTGGGCAAGAGTACCCGCTGTGTGAAGTACCCCTGGAAGGAGGCGCAGCCCG 163  
Db 166 CGTGCCCCGGCAAGAGCTGGCCCAAGCTGGAGTGGTGGCTGTGGCCACGACCTGGCA 225  
Qy 164 CCTAGCCCGCGGCTGTACACCGTGACCTGAGAGCTTCAAGGTGGGCGAGTGGCA 223  
Db 226 CCAAGTACAGATGGTCCACTCCAGTCCGAGACCGGCTCCAGTCCGCGCCCTCCGCC 285  
Qy 224 GCGTGAATGATGACCGCC 241  
Db 286 TCCCCATCAACGGGCTTC 303

RESULT 1:  
LOCUS BU291883 596 bp mRNA linear EST 09-APR-2002  
DEFINITION BU291883 Y. Ogihara unpublished cDNA library, wh\_Sc Triticum  
aestivum cDNA clone whs15e01 5', mRNA sequence.

ACCESSION BU291883

VERSION BU291883.1 GI:20108433

KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

REFERENCE 1 (bases 1 to 596)

AUTHORS Ogihara Y. and Murali K.

TITLE Expressed genes in Triticum aestivum

JOURNAL Unpublished

COMMENT Contact: Tadasu Shin-i  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yatai, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsunagi@res.nig.ac.jp

FEATURES  
Location/Qualifiers

1..596  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="whs15e01"  
/tissue\_type="seed DP10"  
/dev\_stage="feekes' scale 11.3"  
/clone\_lib="Y. Ogihara unpublished cDNA library, wh\_Sc"  
BASE COUNT 118 a 211 c 168 g 94 t 5 others

Query Match 18.6%; Score 49.2; DB 12; Length 586;  
Best Local Similarity 53.0%; Pident No. 1;  
Matches 105; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 44 CCCGACGCGCGGTGACCGCCAGCGGCAAGCCCTGACAGGAGAGAGCTGTGCTACG 103  
Db 144 CACGATTGAGAGTGGCCCCCAAGCCCAACGCTTATGCGGTGGGAAAGGAGCCACT 165  
Qy 104 TGAACCTGGGCAAGAGTACCCGCTGTGTGAAGTACCCCTGGAAGGAGGCGCAGCCCG 163  
Db 224 CGTGCCCCGGCAAGAGCTGGCCCAAGCTGGAGTGGTGGCTGTGGCCACGACCTGGCA 225  
Qy 164 CCTAGCCCGCGGCTGTACACCGTGACCTGAGAGCTTCAAGGTGGGCGAGTGGCA 223  
Db 264 CCAAGTACAGATGGTCCACTCCAGTCCGAGACCGGCTCCAGTCCGCGCCCTCCGCC 285  
Qy 224 GCGTGAATGATGACCGCC 241  
Db 324 TCCCCATCAACGGGCTTC 341

RESULT 1:  
LOCUS BU388416 947 bp mRNA linear EST 27-NOV-2002  
DEFINITION BU388416 603802278F1 CSECHN57 Gallus gallus cDNA clone CHER774120.5', mRNA  
sequence.

ACCESSION BU388416

VERSION BU388416.1 GI:25757452

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Neopteri;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus

REFERENCE 1 (bases 1 to 947)

AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burr, J.N., Bosch, E.,  
Fong, M.T., Tickle, C., Brown, K.R.M., Wilson, S.A. and Hubbard, S.C.

TITLE A Comprehensive Collection of Chicken cDNAs

JOURNAL Curr Biol. 12 (22), 1965-1969 (2002)

MEDLINE 12445392

COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology (UMIST)

PO Box 98, Manchester, M60 1QR, UK  
Tel: 01612008930  
Fax: 01612350405  
Email: Simon.Hubbard@umist.ac.uk

Location/Qualifiers

1..947  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisek"  
/db\_xref="taxon:9932"  
/clone="CHS775125"  
/dev\_stage="16 day e-bryo"  
/lab\_host="DH103"  
/clone\_lib="CSPROCHN57"  
/note="Organ: brain; Vector: pluscript II KS(+); Site: 1;  
Scori, Site 2; NotI. This construct was constructed from a million independent clones. cDNA  
synthesis was initiated using a 3'-tagged primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunt-ended, ligated to NotI adapters, digested with EcoRI  
, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Benaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

BASE COUNT 138 a 260 c 261 g 188 t

Query Match 18.6%; Score 49.2; DB 13; Length 947;  
 Best Local Similarity 53.3%; Pred. No. 1.1;  
 Matches 108; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

DB 11 TGGAGATCAAGCCGACGAGCCGAGTTCACGACCCGAGCGCTGAGCGCCGACGACA 70  
 1 TGGAGAGCAACCCCAACACACAGAGTGGCGGGGTACGGGGTTTGAATCGCGCTCC 60  
 QY 71 AGCCCTACAGCTGAAAGAGAGAGCTGCTAGCTGAGACTGAGGCAACGAGTACCCCTGC 133  
 DB 61 AGCTCATCGTCCGCTGAGAGCAACAGTTTGTGGAAAGGCGCAACTGAACTCGGTGC 120  
 QY 131 TGGTGAAGATCAACCTGAGACGAGCGGACCGGCTACGCGCCGAGCTGTACACCGTGC 190  
 DB 121 AGCGCGGCTGAAAGCTGAAAGAGATTTCGGGAGAGCTTACCCCGCTTCCCGCCATCCAGT 180  
 QY 191 ACCTGAGCAAGCTTCAGG 208  
 DB 181 CCATCCACGTCCTCCAGG 198

RESULT 12  
 B2566686/c 945 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacs2-164\_6521\_Y2 pacs2-164 pseudomonas aeruginosa genomic clone  
 DEFINITION pacs2-164\_6521, genomic survey sequence.  
 ACCESSION B2566686  
 VERSION B2566686.1 GI:27196709  
 KEYWORDS GSS  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 945):  
 Spencer D.H., Raymond C.K., Smith S.E., Sims S.E., Hastings, N.,  
 Burns J.J., Kaul, R. and Olsen, M.V.  
 TITLE Whole-Genome-Sequence Variation among Multiple Isolates of  
 Pseudomonas aeruginosa Library  
 JOURNAL C. Bacteriol., 120(2): in press  
 CONTACT: Chris K. Raymond  
 GENOME CENTER  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel.: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

## FEATURES

source  
 1..945  
 location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164\_6521"  
 /clone\_lib="pacs2-164"  
 /note="clonital isolate 2-164 Whole genomic shotgun library."

BASE COUNT 156 a 307 c 323 g 158 t 2 others  
 ORIGIN

Query Match 18.6%; Score 49.2; DB 29; Length 945;  
 Best Local Similarity 51.4%; Pred. No. 1.1;  
 Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

DB 36 GTTCACCAAGCCGAGCGGCTGAGACCGGCAAGCCCTACAGCTGAAAGAGAGCT 95  
 DB 446 GCTCGGACCCGGTTCAACGTGGCGCCCTGAGAGCAATGTCGGGTGACCTCGAGCG 387  
 QY 96 GTGCTAGTGAAGCTGGGCAAGAGTACCCCGTGCAGATACACCTGAGAGAGG 155  
 DB 386 TGCGCGGAGAGGTGGCACTCGGCTTACGGGCGCAAGAGCCAGTGTGTCTCTGCGGG 327  
 QY 156 CGAGCCGCTAGCGCCCGCGCTGTACACCGTGAAGCTGAGCAAGCTTGAAGGTGAGTCA 215

DB 326 CCACGAGGTGATGTCATTTGGGAGGCAACTGATGCGCTGGCCCAAGGTGATGCGGAC 267  
 QY 216 GTTCGCAAGCTGATGATGACACCGCTTCGCTGCTGTCGCTGC 257  
 DB 266 GGCATGAGCTGAAAGAGACACCGCTGCTTCGAGCGGAC 225

RESULT 13  
 B0412619 974 bp mRNA linear EST 29-NOV-2002  
 LOCUS 603155254F1 CSEGRB05 Gallus gallus cDNA clone G851168x14.5, mRNA  
 DEFINITION sequence.  
 ACCESSION B0412619  
 VERSION B0412619.1 GI:25305290  
 KEYWORDS EST  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 974):  
 Boardman, P. E., Sanz-Bizquerro, J., Overton, I. M., Burt, D. M., Bosch, E.,  
 Fong, W. T., Rickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 PUBLISHED 22315534  
 PubMed 12445192

COMMENT CONTACT: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology (UMIST)  
 PO Box 99, Manchester, M60 1QD, UK  
 Tel: 0161206930  
 Fax: 0161236409  
 Email: Simon.Hubbard@umist.ac.uk

## FEATURES

source  
 1..974  
 location/Qualifiers  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="layer"  
 /db\_xref="taxon:9913"  
 /cctgc="AGGCTGAGK14"  
 /sex="female"  
 /dev\_stage="adult"  
 /ab\_host="CH18"  
 /clone\_lib="CSEGRB05"  
 /note="Organ: ovary; Vector: pB-uescript II KS(+); Site 1:  
 EcoRI; Site 2: NotI; Modification of B-uescript II KS(+)  
 (Stratagene) vector to accommodate cDNA produced with the  
 T-primed protocol (Construction of uni-directionally  
 cloned cDNA libraries from messenger RNA for improved 3'  
 end DNA sequencing by Glenn Fu, et al. U.S. Patent # 5,387  
 624). Cut B-uescript II KS(+) with NotI and EcoRI.  
 Ligate in double stranded adaptor containing BstXI and  
 BstHI sites (5'ggcggtgtagagcccgagatcgcaaaaagj  
 [5'aattccttttcgattcggatcgagcgcaagc])"

BASE COUNT 183 a 292 c 282 g 210 t  
 ORIGIN

Query Match 18.3%; Score 49.4; DB 23; Length 974;  
 Best Local Similarity 55.3%; Pred. No. 1.6;  
 Matches 94; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

DB 3 GTTGAAGATTAAAGCCAGGCGAGGCCAGTTACACCAAGCCGAGCGCTGAGGCGGCAAGG 68  
 DB 540 GGTGAGAGAGAACCCCAACACACAGAGTGGCGGCGTACGAGGTGGATGCTCGGCTC 593  
 QY 69 CAGAGCCCTACAGCTTGAAGAGAGAGTGTGATGAGCAAGCTGAGAGAGATTAACCCGT 128  
 DB 600 CAGCTCATATGCTCTCTGATGAGCAAGATTTTGTGAAAGAGCCCAAGCTGAACTGCT 653  
 QY 129 GCTGTGAAGATACCTGTGAGACAGAGGCGAGCGCTTACGCCCCCGGCGC 178



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OX nucleic - nucleic search, using sw mode!

Run on: November 7, 2003, 12:27:20 Search time: 187 seconds

(without alignments)  
3810.971 Million cell updates/sec

Title: US-09-622-500B-3

Perfect score: 264

Sequence: : atgataaagtgatgacaa.....gcccgtgcccgaagtaa z64

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: N\_Geneseq\_19Jun03.\*  
2: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1980.DAT.\*  
3: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1981.DAT.\*  
4: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1982.DAT.\*  
5: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1983.DAT.\*  
6: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1984.DAT.\*  
7: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1985.DAT.\*  
8: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1986.DAT.\*  
9: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1987.DAT.\*  
10: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1988.DAT.\*  
11: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1989.DAT.\*  
12: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1990.DAT.\*  
13: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1991.DAT.\*  
14: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1992.DAT.\*  
15: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1993.DAT.\*  
16: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1994.DAT.\*  
17: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1995.DAT.\*  
18: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1996.DAT.\*  
19: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1997.DAT.\*  
20: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA1998.DAT.\*  
21: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA2001.DAT.\*  
23: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA2002.DAT.\*  
24: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA2003.DAT.\*  
25: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	100.0	264	AA211645	Coliphage M13 gene
2	147.2	55.8	6971	AA26304	Sequence of phage
3	145.6	55.2	7055	AA26303	Sequence of phage
4	145.6	55.2	7783	AA26302	Sequence of phage
5	145.6	55.2	8233	AA26301	Sequence of phage
6	144	54.5	264	AA211644	Coliphage M13 gene
7	144	54.5	7084	AA251026	Modified filament
8	144	54.5	7294	AA241170	M13142, Synthe

9	144	54.5	7294	13	AA241174	M131421, Synthe
10	144	54.5	7294	15	AA266626	Right-half vector
11	144	54.5	7294	15	AA266630	Right-half vector
12	144	54.5	7317	19	AA266631	M13-based vector M
13	144	54.5	7317	23	AA266633	plasmid M131X11
14	144	54.5	7317	21	AA2931525	Kappa 199t chain
15	144	54.5	7320	21	AA2931525	Left-half vector M
16	144	54.5	7394	13	AA241175	M13BD4, Synthetic
17	144	54.5	7394	15	AA266632	Left-half vector X
18	144	54.5	7409	15	AA266632	Left-half vector M
19	144	54.5	7445	13	AA241172	M131X10, Synthetic
20	144	54.5	7445	15	AA266628	Vector M131X3 for
21	144	54.5	7445	19	AA266630	M13-based vector X
22	144	54.5	7445	20	AA266631	plasmid M131X3
23	144	54.5	7445	22	AA2931524	Heavy chain Rd 11p
24	144	54.5	7445	19	AA266633	M13-based vector M
25	144	54.5	7557	20	AA266633	plasmid M131X11
26	144	54.5	7557	22	AA2931527	Nucleotide sequence
27	144	54.5	7729	19	AA266632	M13-based vector M
28	144	54.5	7729	20	AA266633	Nucleotide sequence
29	144	54.5	7729	21	AA2931525	M13-based vector M
30	144	54.5	8118	15	AA266634	plasmid M131X3
31	144	54.5	8118	20	AA266634	plasmid M131X3
32	144	54.5	8118	21	AA2931528	Nucleotide sequence
33	144	54.5	9532	25	AA2931528	Nucleotide sequence
34	144	54.5	10251	25	AA2931528	Nucleotide sequence
35	142.4	53.9	7330	13	AA241171	M131X2, Synthetic
36	142.4	53.9	9955	19	AA266632	M131X2, Synthetic
37	142.4	53.9	9955	11	AA266632	M131X2, Synthetic
38	142.4	53.9	9955	13	AA266632	M131X2, Synthetic
39	142.4	53.9	7409	13	AA241173	M131X3, Synthetic
40	133.6	50.6	7552	14	AA266632	Sequence of phage
41	132.6	42.7	7302	22	AA266634	M131X3, Synthetic
42	50.4	19.1	803	25	AA266634	Nucleotide sequence
43	50.4	19.1	803	25	AA266634	Nucleotide sequence
44	50.4	19.1	803	25	AA266634	Nucleotide sequence
45	50.4	19.1	803	25	AA266634	Nucleotide sequence

## ALIGNMENTS

RESULT 1  
AA211645  
ID AA211645 standard; DNA: 264 bp.  
XX  
AC AA211645;  
XX  
DT 19-NOV-1999 (first entry)  
XX  
DE Coliphage M13 gene 5 synthetic nucleic acid sequence.  
XX  
KW Genitivirus resistance; ssDNA virus; ssDNA-binding protein; IVV;  
KW Involucrin virus; Coliphage; gene 5 protein; viral infection;  
KW plant virus; plant protection; ss.  
XX  
XX Coliphage sp.  
OS Synthetic.  
XX  
FH Key location/Qualifiers  
FT CDS 1..264  
FT CDS /\*tag= a  
XX  
XX MO3945101-A1.  
XX  
ED 10-SEP-1999.  
XX  
XX 03-MAR-1999; 99WO-US04716.  
XX  
XX 03-MAR-1999; 98US-0076627.  
XX  
XX (SCRI ) SCRIpps RES INST.





```

XX XX Phage vector; fjun.1B; multimeric protein complex; polyphage particle;
KM antibody Fab fragment; enzyme complex; ds.
XX Synthetic.
XX MO9906587-A2.
XX
XX 11-FEB-1999.
XX
XX 03-AUG-1998; 98WO-EP04836.
XX
XX 01-AUG-1997; 97EP-0113319.
XX
XX (MORP-) MORPHOSYS GES PROTEINOPTIMIERUNG MBH.
XX
XX Ge L, Ilag V, Rudert F;
XX
XX MPI; 1999-153811/13.
XX
XX PT Identifying combinations of nucleic acids that encode multimeric
XX PT protein complexes - based on new polyphage particles that display
XX PT both components simultaneously, particularly for selecting antibody
XX PT Fab fragments or enzyme complexes, does not require a recombination
XX PT event.
XX
XX Example 2; Fig 3; 71pp; English.
XX
XX CC The present sequence represents the phage vector fjun.1B, which was
XX CC used in the course of the invention. The specification describes a
XX CC method for identifying combinations of nucleic acids that encode
XX CC multimeric protein complexes. The method is based on polyphage
XX CC particles that display both components simultaneously, particularly
XX CC for selecting antibody Fab fragments or enzyme complexes. In addition,
XX CC the method does not require a recombination event, since polyphages
XX CC co-package both libraries, both members of the multimeric polypeptide
XX CC complex, particularly the light and heavy chains of an Fab fragment,
XX CC can be identified simultaneously.
XX
XX SQ Sequence 7055 BP; 1769 A; 1433 C; 1492 G; 2357 T; 4 other;
XX
XX Query Match: 55.2%; Score 145.6; DB 20; Length 7055;
XX Best Local Similarity 72.0%; Pred. No. 1.2e-19;
XX Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
XX
XX QY 1 ATGATCAAGGTGAGATCAAGCCGACGAGCCGACGACGACGAGCGGCTGAGC 60
XX DB 843 ATGATTAAGTTGAATTAACCGCTCAAGGCAATTACACCGCTTGTGCTTTCT 902
XX QY 61 CCGCAGGCGAAGCCCTACAGCTGAAACGAGCTGTGTACTGAGACTGGCAAGAG 120
XX DB 903 CGTCAGGCGAAGCTTATTCACCTGATGAGAGCTTGTGATGATTTGGTAA 962
XX QY 121 TACCCGCTGCTGTAAGATCAAGCTGAGAGGAGGAGCCGCTACGCGCGGCTG 190
XX DB 963 TATCCGCTGCTGTCAAGATTAAGCTGCAAGAGAGTCAAGAGCTATGCGCTGTG 1022
XX QY 181 TACAGCGTGCACCTGAGAGCTTCAAGCTGAGGCAAGTGTGAGGCTGATGACAGCC 240
XX DB 1023 TACAGCGTGCATGTGCTGCTGCAAAAGTGTGATGCTGATTTGATGATGACCGT 1082
XX QY 241 CTGCGCTGCTGCGCGCCCAAGTAA 264
XX DB 1083 CTGCGCTGCTGCTGCGGCTAAGTAA 1106
XX
XX RESULT 4
XX ID AAX26302 standard; DNA; 7783 BP.
XX AC AAX26302;
XX
XX 21-MAY-1999 (first entry)

```

```

XX XX Sequence of phage vector fjun.1A.
XX DE
XX XX Phage vector; fjun.1A; multimeric protein complex; polyphage particle;
XX KM antibody Fab fragment; enzyme complex; ds.
XX XX Synthetic.
XX OS
XX MO9906587-A2.
XX
XX 11-FEB-1999.
XX
XX 03-AUG-1998; 98WO-EP04836.
XX
XX 01-AUG-1997; 97EP-0113319.
XX
XX (MORP-) MORPHOSYS GES PROTEINOPTIMIERUNG MBH.
XX
XX Ge L, Ilag V, Rudert F;
XX
XX MPI; 1999-153811/13.
XX
XX PT Identifying combinations of nucleic acids that encode multimeric
XX PT protein complexes - based on new polyphage particles that display
XX PT both components simultaneously, particularly for selecting antibody
XX PT Fab fragments or enzyme complexes, does not require a recombination
XX PT event.
XX
XX Example 2; Fig 3; 71pp; English.
XX
XX CC The present sequence represents the phage vector fjun.1A, which was
XX CC used in the course of the invention. The specification describes a
XX CC method for identifying combinations of nucleic acids that encode
XX CC multimeric protein complexes. The method is based on polyphage
XX CC particles that display both components simultaneously, particularly
XX CC for selecting antibody Fab fragments or enzyme complexes. In addition,
XX CC the method does not require a recombination event, since polyphages
XX CC co-package both libraries, both members of the multimeric polypeptide
XX CC complex, particularly the light and heavy chains of an Fab fragment,
XX CC can be identified simultaneously.
XX
XX SQ Sequence 7783 BP; 1920 A; 1649 C; 1667 G; 2543 T; 4 other;
XX
XX Query Match: 55.2%; Score 145.6; DB 20; Length 7783;
XX Best Local Similarity 72.0%; Pred. No. 1.2e-19;
XX Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
XX
XX QY 1 ATGATCAAGGTGAGATCAAGCCGACGAGCCGACGACGACGAGCGGCTGAGC 60
XX DB 843 ATGATTAAGTTGAATTAACCGCTCAAGGCAATTACACCGCTTGTGCTTTCT 902
XX QY 61 CCGCAGGCGAAGCCCTACAGCTGAAACGAGCTGTGTACTGAGACTGGCAAGAG 120
XX DB 903 CGTCAGGCGAAGCTTATTCACCTGATGAGAGCTTGTGATGATTTGGTAA 962
XX QY 121 TACCCGCTGCTGTAAGATCAAGCTGAGAGGAGGAGCCGCTACGCGCGGCTG 180
XX DB 963 TATCCGCTGCTGTCAAGATTAAGCTGCAAGAGAGTCAAGAGCTATGCGCTGTG 1022
XX QY 181 TACAGCGTGCACCTGAGAGCTTCAAGCTGAGGCAAGTGTGAGGCTGATGACAGCC 240
XX DB 1023 TACAGCGTGCATGTGCTGCTGCAAAAGTGTGATGCTGATTTGATGATGACCGT 1082
XX QY 241 CTGCGCTGCTGCGCGCCCAAGTAA 264
XX DB 1083 CTGCGCTGCTGCTGCGGCTAAGTAA 1106
XX
XX RESULT 5
XX ID AAD48845 standard; DNA; 8233 BP.
XX AC AAD48845;
XX
XX 21-MAY-1999 (first entry)

```

XX 07-MAR-2003 (first entry)  
 XX fthi vector DNA.  
 XX Multiple cloning site; screening; pVIII protein; fthi vector; ds.  
 XX Undertified.  
 XX WO200281635-A2.  
 XX 17-OCT-2002.  
 XX 05-APR-2002; 2002MO-US-0720.  
 XX 05-APR-2001; 2001US-281401P.  
 XX (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX (MCINNIS P.  
 XX Gershom JM, Basher D;  
 XX WPI; 2003-058517/05.  
 XX Novel DNA molecule for producing filamentous phages, has filamentous  
 PT phage DNA into which recombinant DNA with multiple cloning site having  
 PT a series of restriction sites that do not appear in phage DNA, is  
 PT inserted.  
 XX  
 XX Disclosure; Page 77-82; 82pp; English.  
 XX  
 XX The invention relates to a DNA molecule comprising a filamentous phage  
 CC DNA into which a recombinant DNA sequence containing a multiple cloning  
 CC site having a series of restriction sites that do not appear in, has been  
 CC inserted. This DNA molecule is useful for expressing recombinant peptide  
 CC as pVIII chimera protein in mosaic bacteriophages. The invention also  
 CC relates to a method for screening peptides of interest. Filamentous  
 CC phage and the library containing it is useful for screening for a  
 CC molecule which binds to a peptide of interest. It is useful for  
 CC expression and display of combinatorial random peptides. The present  
 CC sequence is fthi vector DNA used to illustrate the method of the  
 CC invention.  
 XX  
 XX Sequence 8233 BP; 2026 A; 1651 C; 1780 G; 2776 T; 0 other;  
 SQ  
 Query Match 55.2%; Score 145.6; DB 25; Length 8233;  
 Best Local Similarity 72.0%; Pred. No. 1.2e-19;  
 Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
 QY 1 ATGATCAAGGTGAGATCAAGCCAGCCAGCCAGCTTACACACCCGCGAGCGGTGAGC 60  
 DB 1470 ATGATCAAGGTGAGATCAAGCCAGCCAGCTTACACACCCGCGAGCGGTGAGC 1529  
 QY 61 CGCCAGGGCAAGCCCTACAGCTTACAGCAGCAGCTTGTGATGATTTGGTAAATGAA 120  
 DB 1530 CGTCAGGGCAAGCCCTTACAGCAGCAGCTTGTGATGATTTGGTAAATGAA 1589  
 QY 121 TACCCGCTGCTGATGAAGATCAAGCCAGCAGCAGCCAGCTTACAGCAGCAGCTG 180  
 DB 1590 TACCCGCTGCTGATGAAGATCAAGCCAGCAGCAGCCAGCTTACAGCAGCAGCTG 1645  
 QY 181 TACACCGTGAACCTGAGCAGCTTAAAGCTGGCCAGCTTGGAGCAGCTGATGAGCCG 240  
 DB 1650 TACACCGTGAACCTGAGCAGCTTAAAGCTGGCCAGCTTGGAGCAGCTGATGAGCCG 1709  
 QY 241 CTGGGCTGTGCGCCGCAAGTAA 264  
 DB 1710 CTGGGCTGTGCGCCGCAAGTAA 1733  
 RESULT 6  
 AA21644  
 ID AA21644 standard; DNA; 264 BP.

XX AA21644;  
 XX 19-NOV-1999 (first entry)  
 XX Coliphage M13 gene 5 wild-type nucleic acid sequence.  
 XX Coliphage M13 gene 5 wild-type nucleic acid sequence.  
 XX Gemma virus resistance; scDNA virus; scDNA-binding protein; IVV;  
 XX Inoviridae virus; Coliphage; gene 5 protein; viral infection;  
 XX plant virus; plant protection; ss.  
 XX Coliphage sp.  
 XX  
 XX Key Location/Qualifiers  
 XX CDS 1..264  
 XX /\*tag= a  
 XX  
 XX WO9945101-A1.  
 XX 10 SEP-1999.  
 XX 03-MAR-1999; 59KC-US04716.  
 XX 03-MAR-1998; 58US-C076627.  
 XX (SCRI ) SCRIPPS RES INST.  
 XX Padidan M, Beachy RN, Fauquet CM;  
 XX WPI; 1999-551037/46.  
 XX P-P5DB; AA0339639.  
 XX Producing plants resistant to single stranded DNA virus, using a single  
 FT stranded DNA-binding protein of the inoviridae virus family.  
 FT  
 XX Claim 9; Page 6.; 63pp; English.  
 XX  
 XX The invention relates to a method for producing in a plant gemma virus  
 CC resistance to a single stranded (ss) DNA virus. The method comprises  
 CC introducing a gene capable of expressing a scDNA-binding protein of the  
 CC inoviridae virus (IVV) family into the plant. The IVV family virus is  
 CC selected from inoviruses or pleiotovirus genera where the inovirus genus  
 CC virus is selected from a group consisting of Coliphage, enterobacteria  
 CC phage, Pseudomonas phage, Vibriophage or Xanthomonas phage species. The  
 CC scDNA-binding protein is a Coliphage coat protein or a gene 5 protein  
 CC and especially a Coliphage M13 gene 5 protein. The scDNA-binding protein  
 CC of the IVV family interferes with virus spread during the infection  
 CC process of plant viruses of the scDNA type by inhibiting virus spread,  
 CC the virus infection is reduced and/or blocked, thereby increasing plant  
 CC resistance to the virus infection. The methods can be used for protecting  
 CC plants against e.g. Batya streak virus, Bean yellow dwarf virus, Bromus  
 CC striata mosaic virus, Chickpea chlorotic dwarf virus, Chloris striata  
 CC mosaic virus, Digitaria streak virus, Digitaria striata mosaic virus,  
 CC maize streak virus, Mescanthus streak virus, Tobacco yellow dwarf virus,  
 CC wheat dwarf virus, Beet curly top virus, Horse radish curly top virus,  
 CC tomato leafroll virus, Tomato pseudo-curly top virus, Abutilon mosaic  
 CC virus, Acalypha yellow mosaic virus, cassava mosaic virus, Tomato leaf  
 CC curl virus, Cotton leaf curl virus, Bariana bunchy top virus, Coconut  
 CC foliar decay virus, Faba bean necrotic yellows virus, Milk vetch dwarf  
 CC virus or Subterranean clover stunt virus. The present sequence  
 CC represents the wild-type DNA sequence of Coliphage M13 gene 5 protein.  
 XX  
 XX Sequence 264 BP; 58 A; 61 C; 56 G; 89 T; 0 other;  
 SQ  
 Query Match 54.5%; Score 144; DB 20; Length 264;  
 Best Local Similarity 71.6%; Pred. No. 2.6e-19;  
 Matches 139; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
 QY 1 ATGATCAAGGTGAGATCAAGCCAGCCAGCTTACACACCCGCGAGCGGTGAGC 60  
 DB 1 ATGATCAAGGTGAGATCAAGCCAGCCAGCTTACACACCCGCGAGCGGTGAGC 60  
 QY 61 CGCCAGGGCAAGCCCTACAGCTTACAGCAGCAGCTTGTGATGATTTGGTAAATGAA 120

Db 61 CCGTACAGGCGAAGGCTTATTCACCTGATGAGCAGGCTTGTGATGCTTATGATGATGATA 120  
 QY 121 TACCCGCTGCTGATGAGATCACCCTGACGAGGCGCAGCCGCTTACGCCCGGCTG 180  
 Db 122 TATCCGCTGCTGATGAGATCACCCTGATGAGGCTGACGCCGCTTACGCCCGGCTG 180  
 QY 181 TACACGCTGACCTGAGCAGCTTCAAGGCTGCGGCGAGCTTGGCAGGCTGATGATGAGAGCGC 240  
 Db 181 TACACGCTGATCTGTCTCTTTCACAAAGTGGTCAAGTGGTGGTCTTATGATGATGAGCGT 240  
 QY 241 CTGCGGCTGCTGCGGCGGCAAGTAA 264  
 Db 241 CTGCGGCTGCTGCGGCGGCAAGTAA 264

## RESULT 7

AAZ51028  
 ID AAZ51028 standard; DNA; 7084 BP.

AC AAZ51028;  
 CT 05-JUN-2000 (first entry)  
 XX Modified filamentous phage.  
 DE Modified filamentous phage.  
 XX Modified filamentous phage; polypeptide display; expression vector;  
 KW Major coat protein; MCP; ligand-binding heterodimeric antibody; ligand;  
 KM 65.  
 XX Bacteriophage.  
 CS WO200009715-A1.  
 PN 24-FEB-2000.  
 XX 11-AUG-1999; 99WO-US0207.  
 PF 12-AUG-1998; 98US-0096326.  
 PR 24-DEC-1998; 98US-0198639.  
 XX (Scrip) SCRIPPS RES INST.  
 PA Kang AS.  
 PI WPI; 2000-224350/13.  
 DR New modified filamentous phage vectors, useful for the phage display of  
 PT polypeptides.  
 PS Disclosure; Fig 4; 38pp; English.

XX The present sequence is a modified filamentous phage used for modulation  
 CC of polypeptide display on the surface which permits facile manipulation  
 CC of the valency of display. The gene encoding the polypeptide is fused  
 CC into a synthetic copy of a major coat protein (MCP) which enables  
 CC incorporation into the phage during assembly of the filament. Modified  
 CC filamentous phage expression vector includes a wild type MCP gene; a  
 CC leaky inducible promoter; a synthetic MCP gene; and a directional  
 CC cloning site for receiving a nucleotide insert. The vectors are used  
 CC for the expression of polypeptides such as ligand-binding heterodimeric  
 CC antibodies. They can be used for identifying ligands and for  
 CC characterizing antibodies.

XX Sequence 7084 BP; 1739 A; 1469 C; 1485 G; 2391 T; 0 other;

Query Match 54.5%; Score 144; DB 21; Length 7084;  
 Best local similarity 71.6%; Pred No. 2.5e-19;

Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 AGGATCAAGGTGAGATCAAGCCAGGCGGCTTACGACCGGCGAGCGGCTGAGC 60  
 Db 1712 ATGATTAAGTTGAATTAACCATCTCAAGGCCAATTACACTGCTTGGTGTCTT 1771

QY 61 CCGTACAGGCGAAGGCTTATTCACCTGATGAGCAGGCTTGTGATGCTTATGATGATGATA 120  
 Db 1772 CCGTACAGGCGAAGGCTTATTCACCTGATGAGCAGGCTTGTGATGCTTATGATGATA 180  
 QY 121 TACCCGCTGCTGATGAGATCACCCTGACGAGGCGCAGCCGCTTACGCCCGGCTG 180  
 Db 181 TACCCGCTGCTGATGAGATCACCCTGATGAGGCTGACGCCGCTTACGCCCGGCTG 180  
 QY 181 TACACGCTGACCTGAGCAGCTTCAAGGCTGCGGCGAGCTTGGCAGGCTGATGATGAGAGCGC 240  
 Db 181 TACACGCTGATCTGTCTCTTTCACAAAGTGGTCAAGTGGTGGTCTTATGATGATGAGCGT 240  
 QY 241 CTGCGGCTGCTGCGGCGGCAAGTAA 264  
 Db 241 CTGCGGCTGCTGCGGCGGCAAGTAA 264

## RESULT 8

AAQ24170  
 ID AAQ24170 standard; DNA; 7294 BP.

AC AAQ24170;  
 CT 19-JAN-1993 (first entry)  
 XX M13; cloning vectors; gVIII; M13X42; gIXX40; pseudo-wild type;  
 KW coat protein; RES; 55.  
 XX Synthetic.  
 OS WO9206176-A.  
 PN 16-APR-1992.  
 XX 27-SEP-1991; 91WO-US07141.  
 PF 28-SEP-1990; 90US-0590664.  
 PR (IXSV-) IXSVS -NC.  
 XX Huse MD;  
 PI WPI; 1592-150863/18.

XX Surface expression libraries of randomised peptides;  
 PT competing vectors contg. diverse populations of  
 PT oligonucleotide(s) having desirable bias of random codon  
 PT sequences  
 PS Disclosure; Fig 5; 142pp; English.

XX The sequences given in AAQ24170-2 are vectors which allow the  
 CC expression of random peptides on the surface of M13. The vectors  
 CC are produced by combining separate vectors into a single larger  
 CC vector. This system produces random oligonucleotides functionally  
 CC linked to expression elements and to gVIII. M13X42 is used for  
 CC sense strand oligonucleotide portions. It encodes a pseudo-wild type  
 CC gVIII product. This gene encodes the wild-type M13 gVIII amino acid  
 CC sequence but has been changed at the nucleotide level to reduce  
 CC homo-ologous recombination with the wild-type gVIII contained on the  
 CC same vector. The wild-type gVIII is present to ensure that at least  
 CC some functional, non-fusion coat protein is produced. M13X42 is used  
 CC for anti-sense strand oligonucleotide portions. It contains the  
 CC expression elements for the peptide fusion proteins. A ribosome  
 CC binding site and lacZ promoter/operator elements are present for  
 CC transcription and translation of the peptide fusion proteins. M13X30  
 CC contains a wild-type and pseudo-wild-type gVIII genes and various  
 CC restriction sites for cloning of random peptides.

XX Sequence 7294 BP; 1789 A; 1525 C; 1554 G; 2426 T; 0 other;

```

Query Match Similarity: 54.53; Score 144; DB 13 Length 7294;
Best Local Similarity: 71.68; Pred. No. 2.5e-19;
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGAGATCAAGGCCAGGCCAGGCCAGTTTACACCCCGCAGCCGCTGAGC 60
DB 643 ATGATTAAGTTGAATATTAACCACTCAAGGCCCAATTACTACTGTTCTGCTTCT 902
QY 61 CGCCAGGSCAAGCCCTACAGCCTGAACAGACAGCTGCTCTAGTGGACTTGGCAACAG 122
DB 903 CGTCAAGGCAAGCCTTATTCATCTGAATGGACGCTTTGTAGCTTGATTTGGATATGAA 962
QY 121 TACCCCGTCTCTGTGTAAGATCAACCTCTGACAGAGGAGCCAGCCCTACGCCCCGCTG 182
DB 963 TATCCGCTTCTGTTCAGATTAATCTCTTGATGAAGGTCAAGCCAGCCTATACCGCTGGCTG 1222
QY 181 TACACCCCTGACACTGACAGACGCTTGAAGCTCGGCCAGCTTGGCAGAGCTATATGACAGC 242
DB 1023 TACACCGCTTCAATCTGTCTCTTCAAAAGTTGTGTGGTGGTCCCTTATGATTAACCT 1382
QY 241 CTGCCCTCTGTGCTGCCCGCCAGGTAA 264
DB 1083 CTGCCCTCTGTGCTGCCGTAAAGTAA 1106

RESULT 9
AAQ24174
ID AAQ24174 standard; DNA; 7294 BP.
AC AAQ24174;
XX
XX
XX 19-JAN-1993 (first entry)
XX
XX M13; cloning vectors; gvi11; pseudo-wild type; coat protein; RBS;
XX lacZ; ss.
XX
XX Synthetic.
XX
XX WO9206176-A.
XX
XX 16-APR-1992.
XX
XX 27-SEP-1993; 91WO-US07141.
XX
XX 28-SEP-1990; 90US-C590664.
XX
XX (IXSY-) IXSYS INC.
XX
XX Huse MD;
XX
XX
XX WP1: 1992-150863/18.
XX
XX
XX
XX
XX Surface expression libraries of randomised peptide(s) -
XX comprising vectors contg. diverse populations of
XX oligonucleotide(s) having desirable bias of random codon
XX sequences
XX
XX
XX
XX
XX Disclosure; Page 85-88; 142pp; English.
XX
XX
XX
XX
XX The sequences given in AAQ24174-5 are vectors which allow the expression
XX of random peptides on the surface of M13. The vectors are produced by
XX combining separate vectors into a single larger vector. This system
XX produces random oligonucleotides functionally linked to expression
XX elements and to gvi11. These vectors encode a pseudo-wild type gvi11
XX product. This gene encodes the wild-type M13 gvi11 amino acid sequence
XX but has been changed at the nucleotide level to reduce homocytous
XX recombination with the wild-type gvi11 contained in these vectors.
XX
XX The wild-type gvi11 is present to ensure that at least some functional,
XX non-fusion coat protein is produced. They also contain the expression
XX elements for the peptide fusion proteins. A ribosome binding site and

```

```

CC      LacZ promoter/operator elements are present for transcription and
CC      translation of the peptide fusion proteins. Various restriction
CC      sites are present for the cloning of random peptides.
XX
SQ      Sequence 7294 BP; 1796 A; 1567 C; 1553 G; 2418 T; 0 other;
CY      1 ATGATCAAGGTGAGATCAAGCCGACGCGAGCCCGCACTTCAACACGCGACGCGCTGACG 60
DB      Best local similarity: 71.6%; Pred. No: 2,5e-19;
DB      Matches: 189; Conservative: 0; Mismatches: 75; Indels: 0; Gaps: 0;
CY      2 ATGATCAAGGTGAGATCAAGCCGACGCGAGCCCGCACTTCAACACGCGACGCGCTGACG 60
DB      943 ATGATCAAGGTGAGATCAAGCCGACGCGAGCCCGCACTTCACTCGCTTCGAGTTCCT 902
CY      61 CGCAGCGCGACGCGCTTACGCTGACGCGAGCGAGCGTGGCTACGCGGACCGCGGACGAG 120
DB      903 CGTACGCGCGACGCGCTTATTCACGCGAGCGAGCGCTTTGTTACGAGATTGCGGAAATGAA 962
CY      121 TACCGCGCTGCTGTGAGAGATCAAGCGTGGACGAGCGCGCGCGCTTACCGCGCGCGCTG 190
DB      963 TATCGCGCTGCTGTGAGAGATCAAGCGTGGACGAGCGCGCGCGCTTACCGCGCGCGCTG 1022
CY      181 TACCGCGTGCACCTGAGAGAGCGCTTGAAGTGGCGCGCTTGGAGAGCGCGAGATGACCGG 240
DB      1023 TACCGCGTGCACCTGAGAGAGCGCTTGAAGTGGCGCGCTTGGAGATGAGAGCGCT 1082
CY      241 CTGCGCGCTGCTGAGCGCGCGAGTAA 264
DB      1083 CTGCGCGCTGCTGAGCGCGCGAGTAA 1106
XX
RESULT 10
AA666626
ID      AA666626 standard; DNA; 7294 BP.
XX
AC      AA666626;
XX
CT      25-MAR-2003 (updated);
CT      20-JAN-1995 (first entry);
XX
DE      Right-half vector M131X42.
XX
KM      vector M131X42; right-half vector; randomised oligonucleotides;
KM      surface expression; random peptide; constrained secondary structure;
KM      pseudo wild-type M13 gene VIII; ds.
XX
OS      Synthetic.
XX
WC      MO5411496-A1.
XX
XX      26-MAY-1994.
XX
XX      09-NOV-1993; 93MO-US-0850.
XX      18-NOV-1992; 92US-0978893.
XX      (IXSY-1) IXSYS INC.
XX
XX      Muse WC;
XX
XX      WP1; 1994-181498/22.
XX
PT      Cells expressing oligonucleotide(s) having random codon
PT      sequences are used for producing soluble peptide(s) having a
PT      constrained secondary structure in situ.
XX
XX      Example 1; Page 76-82; 152pp; English.
XX
CC      M13X2 was constructed to harbor the right-half populations of
CC      randomised oligonucleotides. M13mp was the starting vector which
CC      was modified to contain, in addition to the encoded wild-type M13
CC      gene VIII all ready present in the vector, a pseudo-wild-type M13
CC      gene VII; sequence with a stop codon placed between it and an EcoRI-

```

CC SacI cloning site for randomised oligonucleotides; a pair of PstI  
 CC sites to be used for joining with M13X22, the left-hand vector; a  
 CC second stop codon placed on the opposite side of the vector than the  
 CC portion being combined with the left-half vector; and various other  
 CC mutations to remove redundant restriction sites and the N-terminal  
 CC portion of LacZ.  
 CC (Updated on 25-MAR-2003 to correct PK field.)  
 CC XX

Sequence 7294 BP; 1787 A; 1528 C; 1552 G; 2427 T; 0 other;  
 Query Match 54.5%; Score 144; DB 15; Length 7294;  
 Best Local Similarity 71.6%; Pred. No. 2,56-19;  
 Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGGAGATCAAGCCCAAGCCGACCTTACACCCGAGCGGCTAGC 60  
 DB 843 ATGATTAAGTTGAAATTAAACCATCTCAAGCCCAATTACTACGTTCTGCTTCT 902  
 QY 61 CGCCAGGGCAAGCCCTACAGCCTGAACAGAGAGCTGTCTACGTGGACCTGGCAAGAG 120  
 DB 903 CGTCAAGGCAAGCCTTATTCATGAAATGAGCAGCTTTGTACGTGATTTGGTAAATGAA 962  
 QY 121 TACCCGCTCTGTGTAAGATCAACCTTGACAGAGGCGCCGCTTACGCCCTG 180  
 DB 963 TATCGGCTCTGTGTAAGATCACTTTGATGAGGTCAAGCCGCTTACGCCCTGCTG 1022  
 QY 181 TACACCGTGCACCTGAGCAGCTTCAAGGTGGCGCAGTTGGCAGCCTGATGATGACCGC 240  
 DB 1023 TACACCGTGCATCTGCTCTTCAAAAGTGGTGTGCTGCTGCTTACCTATGATTAACCGT 1082  
 QY 241 CTGCGCTCTGTGCTCCGCCCAAGTAA 264  
 DB 1083 CTGCGCTCTGTGCTCCGCTAAGTAA 1106

RESULT 11  
 ID AAG66630 standard; DNA; 7294 BP.  
 XX  
 AC AAG66630;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 23-JAN-1995 (first entry)  
 XX  
 DE Right-half vector M13X421.  
 XX  
 KM vector M13X421; right-half vector; randomised oligonucleotides;  
 KM surface expression; random peptide; constrained secondary structure;  
 KM pseudo wild-type M13 gene VII; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN MO9411436-A1.  
 XX  
 PD 26-MAY-1994.  
 XX  
 PF 09-NOV-1993; 93MO-US10853.  
 XX  
 PR 10-NOV-1992; 92US-0978893.  
 XX  
 PA (IXSY-) IXSYS INC.  
 XX  
 PI Huse WD;  
 XX  
 DR WPI; 1994-183498/22.  
 XX  
 PT Cells expressing oligo:nucleotide(s) having random codon  
 PT sequences - are used for producing soluble peptide(s) having a  
 PT constrained secondary structure in soln.  
 XX  
 PS Example 3; Page 94-97; 152pp; English.  
 XX  
 CC M13X421 is identical to vector M13X42 (AAG66626) except that the

CC amber codon between the EcoRI-SacI cloning site and the pseudo-wild  
 CC type gene VII sequence was removed. This change ensures that all  
 CC expression of the LacZ reporter produces a peptide-gene VII  
 CC fusion protein. The vector was used for the construction of 1992  
 CC half oligonucleotide libraries.  
 CC (Updated on 25-MAR-2003 to correct PK field.)  
 CC XX

Sequence 7294 BP; 1787 A; 1528 C; 1553 G; 2426 T; 0 other;  
 Query Match 54.5%; Score 144; DB 15; Length 7294;  
 Best Local Similarity 71.6%; Pred. No. 2,56-19;  
 Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGGAGATCAAGCCCAAGCCGACCTTACACCCGAGCGGCTAGC 60  
 DB 843 ATGATTAAGTTGAAATTAAACCATCTCAAGCCCAATTACTACGTTCTGCTTCT 902  
 QY 61 CGCCAGGGCAAGCCCTACAGCCTGAACAGAGAGCTGTCTACGTGGACCTGGCAAGAG 120  
 DB 903 CGTCAAGGCAAGCCTTATTCATGAAATGAGCAGCTTTGTACGTGATTTGGTAAATGAA 962  
 QY 121 TACCCGCTCTGTGTAAGATCAACCTTGACAGAGGCGCCGCTTACGCCCTG 180  
 DB 963 TATCGGCTCTGTGTAAGATCACTTTGATGAGGTCAAGCCGCTTACGCCCTGCTG 1022  
 QY 181 TACACCGTGCACCTGAGCAGCTTCAAGGTGGCGCAGTTGGCAGCCTGATGATGACCGC 240  
 DB 1023 TACACCGTGCATCTGCTCTTCAAAAGTGGTGTGCTGCTGCTTACCTATGATTAACCGT 1082  
 QY 241 CTGCGCTCTGTGCTCCGCCCAAGTAA 264  
 DB 1083 CTGCGCTCTGTGCTCCGCTAAGTAA 1106

RESULT 12  
 ID AAV03651 standard; DNA; 7317 BP.  
 XX  
 AC AAV03651;  
 XX  
 DT 02-APR-1998 (first entry)  
 DT 02-APR-1998 (first entry)  
 XX  
 DE M13-based vector M13X42.  
 XX  
 KM Vector construction; T-cell receptor; antibody production; diagnosis;  
 KM heteromeric receptor; immunoglobulin superfamily; therapy; circular;  
 KM M13-based vector; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN US5698426-A.  
 XX  
 PD 16-DEC-1997.  
 XX  
 PF 05-JUN-1995; 95US-0464-36.  
 XX  
 PR 27-SEP-1992; 91US-0767-36.  
 XX  
 PR 28-SEP-1990; 90US-0590215.  
 XX  
 PR 13-SEP-1993; 93US-0120648.  
 XX  
 PR 01-DEC-1994; 94US-0349131.  
 XX  
 PR 05-JUN-1995; 95US-0464-36.  
 XX  
 PA (IXSY-) IXSYS INC.  
 XX  
 PI Huse WD;  
 XX  
 DR WPI; 1996-051489/05.  
 XX  
 PT Construction of heteromeric receptor libraries - by combining  
 PT vectors encoding different immunoglobulin fragments  
 XX  
 PS Example 1; Column 27-34; 57pp; English.  
 XX  
 CC

CC This sequence represents a M3-based vector constructed using the method  
CC of the invention, and the primers shown in AA03629-V03649. The method  
CC of the invention is for constructing a diverse population of vectors  
CC capable of expressing diverse populations of first and second  
CC polypeptides which form a diverse population of heteromeric receptors of  
CC the immunoglobulin superfamily, and comprises: (a) operationally linking  
CC to a first vector a first population of diverse DNA sequences within a  
CC first gene family encoding a diverse population of first polypeptides of  
CC the heteromeric receptor of the immunoglobulin superfamily; the first  
CC vector having two pairs of restriction sites symmetrically oriented about  
CC a cloning site; (b) operationally linking to a second vector a second  
CC population of diverse DNA sequences within a second gene family encoding  
CC a diverse population of second polypeptides of the heteromeric receptor  
CC of the immunoglobulin superfamily; the second vector having two pairs of  
CC restriction sites symmetrically oriented about a cloning site in an  
CC identical orientation to that of the first vector; where polypeptides  
CC encoded by the first and second gene families are known to form  
CC heteromeric receptors of the immunoglobulin superfamily; and (c)  
CC combining the vector sequences produced by steps (a) and (b) under  
CC conditions which allow only the operational combination of the vector  
CC sequences containing the first and second populations of diverse DNA  
CC sequences. The method is used to produce diverse populations of  
CC antibodies or T-cell receptors for diagnostic or therapeutic purposes.

SC Sequence 7317 BP; 1787 A; 1555 C; 1547 G; 2428 T; 0 other;

Query Match 54.5%; Score 144; DB 19; Length 7317;  
Best Local Similarity 71.6%; Pred. No. 2.5e-19;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTCAACCCGCGAGCGGCTGAGC 60  
DB 843 ATGATCAAGTGGAGATCAAGCCAGCCAGCTTCAACCCGCGAGCGGCTGAGC 60  
Dc 903 CGTACGGGCGAAGCTTATGACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
QY 121 TACCCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 180  
Db 963 TATCCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 1022  
QY 181 TACACCGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 240  
Db 1223 TACACCGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 1082  
QY 241 CTGCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 264  
Db 1083 CTGCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 1106

RESULT 13  
AA016953  
ID AA016953 standard; DNA: 7317 BP.

AC AA016953;

XX 11-MAY-1999 (first entry);

DE Plasmid M31X11.

XX Heteromeric receptor; immunoglobulin superfamily; plasmid; primer; PCR;

KW bacteriophage; fusion protein; amplification; heavy chain; light chain;

KW immune system; diagnosis; cyclic; circular; ds.

XX Synthetic.

XX US871974-A.

XX 16-FEB-1999.

XX 02-DEC-1994; 94US-0349131.

PR 27-SEP-1994; 94US-0767136.  
PR 28-SEP-1990; 90US-0590219.  
PR 15-SEP-1993; 93US-0120648.  
PR 02-DEC-1994; 94US-0149131.  
XX (IXSY-) IXSY- INC.  
XX Huse KD;  
XX WPI: 1993-166647/14.  
XX New surface expression libraries expressing heteromeric receptors  
XX comprising cells containing vectors containing combinations of DNA  
XX sequences encoding first and second polypeptides  
XX Example 1: Fig 3A-C; 58pp; English.  
XX The invention relates to the expression of heteromeric receptor  
XX proteins, e.g. from an immunoglobulin (Ig) superfamily, in cells  
XX containing the heteromeric receptor genes on a single plasmid.  
XX Especially mentioned, the cell may be a bacteriophage, where the receptor  
XX protein are expressed as fusion proteins with the surface protein GVI.  
XX This sequence represents the plasmid M31X11 which is used for expression  
XX of the light chain proteins. The plasmid is constructed using primers  
XX AA016953-016952. The methods can be used to generate diverse populations  
XX of heteromeric receptors which mimic the natural immune system and can be  
XX used for diagnostic and therapeutic purposes.

SC Sequence 7317 BP; 1787 A; 1555 C; 1547 G; 2428 T; 0 other;  
Query Match 54.5%; Score 144; DB 20; Length 7317;  
Best Local Similarity 71.6%; Pred. No. 2.5e-19;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTCAACCCGCGAGCGGCTGAGC 60  
Db 843 ATGATCAAGTGGAGATCAAGCCAGCCAGCTTCAACCCGCGAGCGGCTGAGC 60  
Dc 903 CGTACGGGCGAAGCTTATGACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
QY 121 TACCCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 180  
Db 963 TATCCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 1022  
QY 181 TACACCGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 240  
Db 1223 TACACCGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 1082  
QY 241 CTGCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 264  
Db 1083 CTGCGCTGGTGGAGATCAAGCTTCAACCCGCGAGCGGCTGAGC 1106

RESULT 14  
AA016952  
ID AA016952 standard; DNA: 7317 BP.

AC AA016952;

XX 25-MAY-2000 (first entry)

XX Kappa light chain library nucleotide sequence of M31X11.

XX Bacteriophage M3 vector; prokaryotic cell; heteromeric receptor;

KW antibody; immune system; filamentous bacteriophage; cloning;

XX screening; coexpression; ds.

XX Mus sp.

XX Bacteriophage M3.

XX Synthetic.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Inc.

OW nucleic - nucleic search, using sw model

Run on: November 7, 2003, 12:56:12 ; Search time 194 Seconds

(without alignments)  
4342.314 Million cell updates/sec

Title: US-09-622-500b-3  
Perfect score: 264  
Sequence: 1 atgacacagtgagagatcaaa.....gcctggtccgcacagatca 264

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 214354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 5  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

```

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US04_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US05_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US6c_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US6c_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	54.5	7083	10	US-09-995-396-1
2	144	54.5	7294	9	US-09-727-311-1
3	144	54.5	7294	9	US-09-727-311-2
4	144	54.5	7320	9	US-09-727-311-5
5	144	54.5	7394	9	US-09-727-311-6
6	144	54.5	7409	9	US-09-727-311-4
7	144	54.5	7445	9	US-09-727-311-3
8	144	54.5	8195	12	US-10-378-557-57
9	133.6	50.6	8171	12	US-10-378-557-58
10	54.2	20.5	2133	14	US-10-156-761-997
11	54.2	20.5	9025608	14	US-10-156-761-1
12	50.4	19.1	759	12	US-10-329-386-2795
13	46.6	17.7	972	12	US-10-259-165-321
14	46.4	17.6	1242	14	US-10-156-761-5818
15	46.4	17.6	9025608	14	US-10-156-761-1
16	45.6	17.3	2010	12	US-10-414-637-9

17	45.6	17.3	2010	13	US-10-329-386-2795	Sequence 3, Appl
18	45.4	17.2	732	14	US-10-156-761-608	Sequence 6808, Ap
19	45.4	17.2	1782	14	US-10-156-761-5623	Sequence 5623, Ap
20	45.2	17.1	1479	12	US-09-367-464-67	Sequence 67, Appl
21	45	17.0	1458	14	US-10-156-761-6684	Sequence 6684, Ap
22	44.4	16.8	1479	12	US-09-367-464-63	Sequence 63, Appl
23	44.4	16.8	1506	12	US-09-367-464-66	Sequence 66, Appl
24	44.4	16.8	2742	12	US-10-190-435-20	Sequence 20, Appl
25	44.4	16.8	2742	12	US-10-190-435-57	Sequence 57, Appl
26	44.4	16.8	2742	12	US-10-190-305A-15	Sequence 15, Appl
27	44.4	16.8	2742	12	US-10-190-305A-81	Sequence 81, Appl
28	44.4	16.8	2739	12	US-10-241-009-18	Sequence 18, Appl
29	44.4	16.8	2739	12	US-10-241-009-53	Sequence 53, Appl
30	44.4	16.8	2739	12	US-10-190-435B-18	Sequence 18, Appl
31	44.4	16.8	2739	12	US-10-190-435B-53	Sequence 53, Appl
32	44.4	16.8	2739	12	US-10-190-305A-11	Sequence 11, Appl
33	44.4	16.8	2739	12	US-10-190-305A-74	Sequence 74, Appl
34	44.4	16.8	3162	12	US-10-190-435-18	Sequence 18, Appl
35	44.4	16.8	3205	12	US-10-241-009-17	Sequence 17, Appl
36	44.4	16.8	3205	12	US-10-190-435B-17	Sequence 17, Appl
37	44.4	16.8	3482	12	US-10-241-009-16	Sequence 16, Appl
38	44.4	16.8	3482	12	US-10-241-009-15	Sequence 15, Appl
39	44.4	16.8	3482	12	US-10-190-435B-15	Sequence 15, Appl
40	44.4	16.8	3531	12	US-10-190-435-13	Sequence 13, Appl
41	44.4	16.8	3537	12	US-10-190-435-14	Sequence 14, Appl
42	44.4	16.8	3537	12	US-10-190-435-15	Sequence 15, Appl
43	44.4	16.8	3564	12	US-10-241-009-12	Sequence 12, Appl
44	44.4	16.8	3564	12	US-10-241-009-14	Sequence 14, Appl
45	44.4	16.8	3564	12	US-10-190-435B-13	Sequence 13, Appl

## ALIGNMENTS

```

RESULT:
US-09-995-396-1
Sequence 1, Application US/09995396
Patent No. US20020446827A1
GENERAL INFORMATION:
APPLICANT: Kary, Argray S.
TITLE OF INVENTION: MODULATION OF POLYMERASE DISPLAY ON NUCLEIC
FILE REFERENCE: SCR22215
CURRENT APPLICATION NUMBER: US/09/995,396
CURRENT FILING DATE: 2001-11-27
PRIOR APPLICATION NUMBER: 09/552,834
PRIOR FILING DATE: 2001-04-09
PRIC APPLICATION NUMBER: 60/096,326
PRIOR FILING DATE: 1998-08-12
PRIOR APPLICATION NUMBER: 09/98,939
PRIOR FILING DATE: 1998-12-24
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH 7083
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: a modified
OTHER INFORMATION: filamentous phage
US-09-995-396-1

```

Query Match: 54.5% Score 144; DB 10; Length 7083;  
Best Local Similarity 71.6%; Pred. No. 1.7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

DB 1771 ATATTAAGGTGAGATCAAGCCACCCAGGCGGCACTTACCAACCCGCGCGGTGAC 60  
1771 ATATTAAGGTGAGATCAAGCCACCCAGGCGGCACTTACCAACCCGCGGTGAC 60

DB 61 CGCAGGCGCAGCGCTTACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 120  
1771 CGCAGGCGCAGCGCTTACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 120

DB 1771 CGCAGGCGCAGCGCTTACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 180



QY 121 TACCCGCTGCTGGAGATACCCCTGGACGAGGCGCCAGCTTACCCCTCCCGGCTG 180  
DB 1831 TATCCCGTCTTGTGACAGATTAAGATGAGAGCTGAGAGGCTATGGGCTGGCTTG 1892  
QY 181 TACACCGTGACCTGGAGAGCTTCAAGTGGCGAGTGGCGAGCTATGATGAGAGCGC 240  
DB 1891 TACACCGTTCATCTGCTCTTTGAAAGTGGTGAGTGGCTCCCTATGATGAGCGCT 1950  
QY 241 CTGCGCCTGTGCTCCCGCCAGTAA 264  
DB 1951 CTGCGCCTGTGCTCCCGCCAGTAA 1974

## RESULT 2

US-09-727-311-1  
Sequence 1, Application US/39727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William E.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
NUMBER OF SEQUENCES: 61  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 200C  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 9072  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7294 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-1

Query Match 54.5%; Score 144; DB 9; Length 7294;  
Best Local Similarity 71.6%; Pred. No. 1,7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGCTGAGATCAAGCCAGCCAGCCAGCTTCAAGCCCGGAGCGCTGAGC 50  
DB 843 ATGATTAAG-TGAATTAACCATCTCAAGCCCAATTACTCTTCTTGCTGTTCT 902  
QY 61 CGCCAGGCGAAGCCCTACAGCCTTAAGCAGAGCTGCTACGTGAGACTGGGCAAGAG 120  
DB 963 CGTCAGGCGAAGCCCTATTCACTGAAATAGACAGCTTTGTACGTTGATTTGGGTAAGAA 962  
QY 121 TACCCGCTGCTGGAGATCAAGCTGAGAGAGGCGCCAGCCGCTTACGCCCCGCTG 180  
DB 963 TATCCGCTTCTTGTGACAGATTAAGTCAAGAGCTTGAAGAGTGAAGAGCTTGGCTG 1822

QY 121 TACACCGTGACCTGGAGAGCTTCAAGTGGGCGAGCTGCGAGCCCTATGATGAGCGC 240  
DB 1823 TACACCGTTCATCTGCTCTTTGAAAGTGGTGAGTGGCTCCCTATGATGAGCGCT 1982  
QY 241 CTGCGCCTGTGCTCCCGCCAGTAA 264  
DB 1951 CTGCGCCTGTGCTCCCGCCAGTAA 1974

## RESULT 3

US-09-727-311-5  
Sequence 5, Application US/39727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William E.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
NUMBER OF SEQUENCES: 61  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 200C  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 9072  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7294 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-5

Query Match 54.5%; Score 144; DB 9; Length 7294;  
Best Local Similarity 71.6%; Pred. No. 1,7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGCTGAGATCAAGCCAGCCAGCCAGCTTCAAGCCCGGAGCGCTGAGC 50  
DB 843 ATGATTAAG-TGAATTAACCATCTCAAGCCCAATTACTCTTCTTGCTGTTCT 902  
QY 61 CGCCAGGCGAAGCCCTACAGCCTTAAGCAGAGCTGCTACGTGAGACTGGGCAAGAG 120  
DB 963 CGTCAGGCGAAGCCCTATTCACTGAAATAGACAGCTTTGTACGTTGATTTGGGTAAGAA 962  
QY 121 TACCCGCTGCTGGAGATCAAGCTGAGAGAGGCGCCAGCCGCTTACGCCCCGCTG 180  
DB 963 TATCCGCTTCTTGTGACAGATTAAGTCAAGAGCTTGAAGAGTGAAGAGCTTGGCTG 1822  
QY 121 TACACCGTGACCTGGAGAGCTTCAAGTGGGCGAGTGGCGAGCTATGATGAGAGCGC 240  
DB 1823 TACACCGTTCATCTGCTCTTTGAAAGTGGTGAGTGGCTCCCTATGATGAGCGCT 1982  
QY 241 CTGCGCCTGTGCTCCCGCCAGTAA 264

Db 1083 CTGCGCCCTGCTCCGCGCTAAGTAA 1106

## RESULT 4

US-09-727-311-2  
Sequence 2, Application US/09727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2200  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentia Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION/DOCKET NUMBER: 31,815  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7320 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-2

Query Match 54.5%; Score 144; DB 9; Length 7320;  
Best Local Similarity 71.6%; Pred. No. 1,7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGAGATCAAGCCGACGAGCCGAGCTTACGACCCGAGCGGCGTGAAG 60  
Db 843 ATGATTAAGTTGAATTAACCATCTCAAGCCCAATTACTACTGCTTCTGCTTTCT 902  
QY 61 CGCGAGGCGAAGCCCTACAGCCCTGAAGCGAGCGCTGCTGCTGAGCTGGGCAAGAG 120  
Db 903 CGTACAGGCGAAGCCCTATTCACGATGAGCGAGCGCTTGTACGTGATTTGGGTAATGA 962  
QY 121 TACCCCGTGTGCTGAGATCAACCCCTGAGCGAGGCGACCCGCTACGCGCCCGGCTG 180  
Db 963 TATCCGCTTCTTCAAGATTAATCTGTGATGAAGGACGACCGCTAAGCGCTGCTG 1022  
QY 181 TACACCGTGCACCTGAGCGAGCTTCAAGGTGCGCGAGCTTGGCAGGCTGATGATGCAACGCG 240  
Db 1023 TACACCGTGCATCTGCTCTTCAAGGTGCGCGAGCTTGGCAGGCTGATGATGCAACGCT 1082  
QY 241 CTGCGCCTGTGCGCGCGCAAGTAA 264  
Db 1083 CTGCGCCTGTGCGCGCGCAAGTAA 1106

## RESULT 5

US-09-727-311-6  
Sequence 6, Application US/09727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentia Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION/DOCKET NUMBER: 31,815  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ. ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7394 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-6

Query Match 54.5%; Score 144; DB 9; Length 7394;  
Best Local Similarity 71.6%; Pred. No. 1,7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGAGATCAAGCCGACGAGCCGAGCTTACGACCCGAGCGGCGTGAAG 60  
Db 843 ATGATTAAGTTGAATTAACCATCTCAAGCCCAATTACTACTGCTTCTGCTTTCT 902  
QY 61 CGCGAGGCGAAGCCCTACAGCCCTGAAGCGAGCGCTGCTGCTGAGCTGGGCAAGAG 120  
Db 903 CGTACAGGCGAAGCCCTATTCACGATGAGCGAGCGCTTGTACGTGATTTGGGTAATGA 962  
QY 121 TACCCCGTGTGCTGAGATCAACCCCTGAGCGAGGCGACCCGCTACGCGCCCGGCTG 180  
Db 963 TATCCGCTTCTTCAAGATTAATCTGTGATGAAGGACGACCGCTAAGCGCTGCTG 1022  
QY 181 TACACCGTGCACCTGAGCGAGCTTCAAGGTGCGCGAGCTTGGCAGGCTGATGATGCAACGCG 240  
Db 1023 TACACCGTGCATCTGCTCTTCAAGGTGCGCGAGCTTGGCAGGCTGATGATGCAACGCT 1082  
QY 241 CTGCGCCTGTGCGCGCGCAAGTAA 264  
Db 1083 CTGCGCCTGTGCGCGCGCAAGTAA 1106

## RESULT 6

US-09-727-311-4  
Sequence 4, Application US/09727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.





CURRENT FILING DATE: 2002-05-29  
PRIORITY APPLICATION NUMBER: JP 2001-204689  
PRIORITY FILING DATE: 2002-05-30  
PRIORITY APPLICATION NUMBER: JP 2001-272697  
PRIORITY FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1  
LENGTH: 9025608  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4187715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 20.5% Score 54.2; DB 14; Length 9025608;  
Best Local Similarity 54.3%; Pred. No. 2, 1e-05;  
Matches 129; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 15 GATCAAGCCGACGAGCCGACGCTTACACACCCGACGCGAGTACGCGGACGAGGACG 74  
DB 1276169 GATCTACACAGCCGCGCCGACGAGGCGGACGCGCATCTGAGAGACCACTCCGACACCTC 1276229  
QY 75 CTACAGCCTTACACGACGAGCTGTGCTACGTGACCTGAGCTGAGAGTACGCGGCTGCTGCT 134  
DB 1276229 CACTGCTGACGCGCCGCGTGGGCGACACGAGGCTAGCGGACCTACATGCTGCTGCT 1276289  
QY 135 GAAGATACCCCTGACACGAGGAGCA---GCCCCCTTACGCTCCCGGCGCTGTACACCTGTGA 191  
DB 1276289 GGAGACCGAGATGACCTCGGAGGACGTGCTGCGACGACCCGACGCGACCTGCTCA 1276349  
QY 192 CCGAGCAGCTTACAGGTGCGGCGAGCTTGGGACGCGTGTATGAGCGGCTGCGG 246  
DB 1276349 CGAGCGACACTGCGCGGACGAGACCGGACCGGCTTACCGGCTGCTGAGCGGCGCTC 1276403

RESULT 12  
US-10-029-386-22785  
Sequence 22785, Application US/10029386  
Publication No. US2003019470A4;  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEWICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 22785  
LENGTH: 759  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004737.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4  
OTHER INFORMATION: EST HUMAN HIT: AW69906.1, EVALU 0.00e+00  
OTHER INFORMATION: SWISSPROT HIT: Q28142, EVALU 3.10e-01  
OTHER INFORMATION: KJ HIT: AF006529.1, EVALU 1.30e-08  
US-10-029-386-22785

Query Match 19.1% Score 50.4; DB 12; Length 759;  
Best Local Similarity 51.3%; Pred. No. 0.00022;  
Matches 117; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 34 CAGTTACACCGCCGAGCGGCTGAGCGGCGGACGAGGCGCTTACACGCTGAAAGAGAG 93  
DB 439 CCGTTCCACACACACAGCTGCGCGGCTGCGCGGCAACATGTGCAAGACAAAGGTGCTC 498

QY 34 CTGTGCTACGTGACGCTGAGGCAAGAGTACCCGCTGCTGTGAGAGATACCTGAGAGAG 153  
DB 499 AGCGACCTACCTGCGCGGAGCGGCGGACGAGCGGCTGAGAGAGAGAGAGAGAGAG 558  
QY 154 GCGGACGCGGCTTACGCGCGGCGGCTTACACGCTGAGACCTGAGAGAGCTTGMAGAGAG 213  
DB 559 GCGACGCGGCGGACGAGCTTGTGTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 618  
QY 214 CAGTTGAGCAGCTGATGATGACGAGCGGCTGAGCTGAGTGTGAGAGAGAGAG 261  
DB 619 CCGCGCGGCGGCTTACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666

RESULT 13  
US-10-259-165-321  
Sequence 321, Application US/12259165  
Publication No. US2003013588A1  
GENERAL INFORMATION:  
APPLICANT: Zhu, Tong  
APPLICANT: Wang, Xun  
APPLICANT: Chang, Hui-song  
APPLICANT: Briggs, Steven P.  
APPLICANT: Cooper, Bret  
APPLICANT: Glazedrock, Jane  
APPLICANT: Goff, Stephen A.  
APPLICANT: Katagiri, Fumiyaki  
APPLICANT: Krebs, Joel  
APPLICANT: Moughamer, Todd  
APPLICANT: Prevart, Nicholas  
APPLICANT: Rieke, Darrel  
TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENC  
FILE REFERENCE: 72030-NP  
CURRENT APPLICATION NUMBER: US/10/259,165  
CURRENT FILING DATE: 2002-09-26  
PRIORITY APPLICATION NUMBER: US 60/370,620  
PRIORITY FILING DATE: 2002-04-04  
PRIORITY APPLICATION NUMBER: US 60/368,387  
PRIORITY FILING DATE: 2002-03-27  
PRIORITY APPLICATION NUMBER: US 60/325,277  
PRIORITY FILING DATE: 2001-09-26  
NUMBER OF SEQ ID NOS: 782  
SOFTWARE: PatentList v1 version 3.0.4 (C) 2001 Syngenta  
SEQ ID NO 321  
LENGTH: 972  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (135)..  
OTHER INFORMATION: y = cytosine or thymine  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (139)..  
OTHER INFORMATION: m = adenine or cytosine  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (156)..  
OTHER INFORMATION: y = cytosine or thymine  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (169)..  
OTHER INFORMATION: x = adenine or guanine  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (197)..  
OTHER INFORMATION: x = guanine or thymine  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (199)..  
OTHER INFORMATION: x = adenine or guanine  
US-10-259-165-321

Query Match 17.7% Score 46.6; DB 12; Length 972;

Best Local Similarity 54.1%; Pred. No. 0.0023;  
Matches 113; Conservative 2; Mismatches 91; Indels 3; Gaps 1;

```
QY 4 ATCAAGTGTAGATATCAAGCCAGCCAGCCAGCTTACACCCGAGCGGCTAGCGGC 63
DB 196 AAGCGCGCGTGGCCCACTGCGCAACACCCCTCGCGCGCGGAGCTGGGTGGGTAG 255
QY 64 CAGCGAAGCCCTACAGCTGAAAGAGAGCTGTCTGCTAGCTGAGAGAGAGAG 122
DB 256 AAGGTTCAGGCTACTGCGCGAGAGCTATCATCTGGGTACATGCGGTGGGAGAGAG 215
QY 123 --CCGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 316 GCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 375
QY 181 TACAGCTGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 209
DB 376 TCGCGCGCGCTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
```

## RESULT 14

```
US-10-156-761-5818
/ Sequence 5818, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 5818
/ LENGTH: 1242
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1242);
US-10-156-761-5818
```

Query Match 17.6%; Score 46.4; DB 14; Length 1242;  
Best Local Similarity 53.3%; Pred. No. 0.0026;  
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
QY 39 CACCAACCCGAGCGGCTGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 58
DB 1053 CAACCCGAGAGCTGCGCTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1112
QY 99 CTACGTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 158
DB 1113 CGTGTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1172
QY 159 GCGCGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 216
DB 1173 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2232
QY 219 CGGC 222
DB 1233 GAGC 1236
```

## RESULT 15

US-10-156-761-1/c

Sequence 1: Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:

```
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 1
/ LENGTH: 9025608
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (418715);
/ OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

Query Match 17.6%; Score 46.4; DB 14; Length 9025608;  
Best Local Similarity 53.3%; Pred. No. 0.0026;  
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
QY 39 CACCAACCCGAGCGGCTGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 98
DB 7048061 CAACCCGAGAGCTGCGCTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAG 7048062
QY 99 CTACGTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 158
DB 7048061 CGTGTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7047942
QY 159 GCGCGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 218
DB 7047941 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7047892
QY 219 CGGC 222
DB 7047891 GAGC 7047878
```

Search completed: November 7, 2003, 15:56:56  
Cdb time: 219 secs









Matches 189: Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATGATCAAGGTGAGATCAAGCCAGCCAGCCAGCTTACACACCCGACGCGGCTGAGC 60  
DB 843 ATGATTAAGTTTAAATTAAACCATCTCAAGCCCAATTACTACTGCTTCTGCTTCT 902  
QY 61 CGCCAGGCGAAGCCCTTACAGCCCTGAGAGAGAGTGTGCTAGGAGCTTGGGAAAGAG 120  
DB 903 CGTCAGGCGAAGCCCTTATTCAGTAATGAGACAGCTTGTAGCTTATTTGGATATGAA 962  
QY 121 TACCCGCTGCTGAGTGAAGATCAACCTGACAGAGGCTCAGCCGCTACGCGCGGCTG 180  
DB 963 TATCCGCTTCTTCAAGATTAATCTTGAATGAGGTCAAGCCAGCCATGCGCTGCTCG 1022  
QY 181 TACACGCTGACCTGAGACGCTTCAAGGTGCGAGCTTGGGAGCTGAATGACAGCC 240  
DB 1023 TACACGCTTATCTGCTCTTTCCTTTCCTTCAAGTGTGAGCTTCCCTTATGATGACCGT 1082  
QY 241 CTGCGCCTGTGCGCGGCAAGTAA 264  
DB 1093 CTGCGCCTGCTTCCGCTAAGTAA 1106

RESULT 7  
US-08-367-685-5  
Sequence 5, Application US/08367685  
Patent No. 6258530  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Rueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version 4.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/367,685  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,494  
FILING DATE:  
APPLICATION NUMBER: US/07/757,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 9072  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7294 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-08-367-685-5

Query Match 54.5%; Score 144; DB 3; Length 7294;  
Best Local Similarity 71.6%; Pred. No. 1,4e-21;  
Matches 189: Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATGATCAAGGTGAGATCAAGCCAGCCAGCCAGCTTACACACCCGACGCGGCTGAGC 60

DB 843 ATGATTAAGTTTAAATTAAACCATCTCAAGCCCAATTACTACTGCTTCTGCTTCT 902  
QY 61 CGCCAGGCGAAGCCCTTACAGCCCTGAGAGAGAGTGTGCTAGGAGCTTGGGAAAGAG 120  
DB 903 CGTCAGGCGAAGCCCTTATTCAGTAATGAGACAGCTTGTAGCTTATTTGGATATGAA 962  
QY 121 TACCCGCTGCTGAGTGAAGATCAACCTGACAGAGGCTCAGCCGCTACGCGCGGCTG 180  
DB 963 TATCCGCTTCTTCAAGATTAATCTTGAATGAGGTCAAGCCAGCCATGCGCTGCTCG 1022  
QY 181 TACACGCTGACCTGAGACGCTTCAAGGTGCGAGCTTGGGAGCTGAATGACAGCC 240  
DB 1023 TACACGCTTATCTGCTCTTTCCTTTCCTTTCCTTCAAGTGTGAGCTTCCCTTATGATGACCGT 1082  
QY 241 CTGCGCCTGTGCGCGGCAAGTAA 264  
DB 1093 CTGCGCCTGCTTCCGCTAAGTAA 1106

RESULT 8  
PCT-US91-07141-2  
Sequence 1, Application PCT/US9107141  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Rueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version 4.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/07141  
FILING DATE: 19910927  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 9072  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7294 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: both  
TOPOLOGY: circular  
PCT-US91-07141-2

Query Match 54.5%; Score 144; DB 5; Length 7294;  
Best Local Similarity 71.6%; Pred. No. 1,4e-21;  
Matches 189: Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATGATCAAGGTGAGATCAAGCCAGCCAGCCAGCTTACACACCCGACGCGGCTGAGC 60  
DB 843 ATGATTAAGTTTAAATTAAACCATCTCAAGCCCAATTACTACTGCTTCTGCTTCT 902  
QY 61 CGCCAGGCGAAGCCCTTACAGCCCTGAGAGAGAGTGTGCTAGGAGCTTGGGAAAGAG 120  
DB 903 CGTCAGGCGAAGCCCTTATTCAGTAATGAGACAGCTTGTAGCTTATTTGGATATGAA 962  
QY 121 TACCCGCTGCTGAGTGAAGATCAACCTGACAGAGGCTCAGCCGCTACGCGCGGCTG 180

```
Db 963 TATCGGCTTCTGTGTAAGATTAAGTATGTAAGGTCAAGGCAAGGCTATGCGGCTGCTG 1022
Qy 181 TACACCGTGCACCTGAGAGAGCTTCAAGGTCGCGCAAGTTCGGGCAAGCTGATGATGACAGC 240
Db 1023 TACACCGTGCATCTGCTCTCTTCAAAAGTTGCTGAGAGTTGGTTCCCTTATGATGACCGT 1082
Qy 241 CTGGCGCTGCTGCGCGCGCAAGTAA 264
Db 1083 CTGGCGCTGCTGCGCGCTAAGTAA 1106

RESULT 9
PCT-US91-07141-5
; Sequence 5, Application PC/TUS9107141
; GENERAL INFORMATION:
; APPLICANT: Huse, William D.
; TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: United States
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/07141
; FILING DATE: 19910927
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P31 9072
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7294 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; PCT-US91-07141-5

Query Match 54.5%; Score 144; Db 5; Length 7294;
Best Local Similarity 71.6%; Pred. No. 1.4e-21;
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1 ATGATCAAGGTGAGATCAAGGCCAGCGCCAGTTCACACCGCGCGCGCTGAC 60
Db 843 ATGATTAAGTTAAATTAAACATCTCAAGCCCAATTAAGTCTGTTCTTCT 902
Qy 61 CGCCAGGGCAAGGCTTACAGCTGAGAGAGAGCTGCTACCTGGAGCTGGGCAAGAG 120
Db 903 CGTCAGGGCAAGGCTTATTCATGATGAGAGAGCTTGTACGTTGATTGGGTAAAGAA 962
Qy 121 TACCGCGTGTGTGTAAGATCAAGCTTGAGAGAGGCGAGCGCGCTAAGCGCGCGCTG 180
Db 963 TATCGGCTTCTGTCAAGATTAAGTCTTGAAGAGTCAAGCCAGGCTATGCGGCTGATG 1022
Qy 181 TACACCGTGCACCTGAGAGAGCTTCAAGGTGCGGCAAGTTCGGGCAAGCTGATGATGACAGC 240
Db 1023 TACACCGTGCATCTGCTCTCTTCAAAAGTTGCTGAGAGTTGGTTCCCTTATGATGACCGT 1082
Qy 241 CTGGCGCTGCTGCGCGCGCAAGTAA 264
Db 1083 CTGGCGCTGCTGCGCGCTAAGTAA 1106
```

```
RESULT 10
US-08-464-136-2
; Sequence 2, Application US/08464136
; Patent No. 5698426
; GENERAL INFORMATION:
; APPLICANT: HUSE, WILLIAM D.
; TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESS: PRETTY, SCHROEDER, BRUEGEMANN & CLARK
; STREET: 444 S. FLOWER STREET, SUITE 200
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P31 8882
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7317 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; US-08-464-136-2

Query Match 54.5%; Score 144; Db 1; Length 7317;
Best Local Similarity 71.6%; Pred. No. 1.4e-21;
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1 ATGATCAAGGTGAGATCAAGGCCAGCGCCAGTTCACACCGCGCGCGCTGAC 60
Db 843 ATGATTAAGTTAAATTAAACATCTCAAGCCCAATTAAGTCTGTTCTTCT 902
Qy 61 CGCCAGGGCAAGGCTTACAGCTGAGAGAGAGCTGCTACCTGGAGCTGGGCAAGAG 120
Db 903 CGTCAGGGCAAGGCTTATTCATGATGAGAGAGCTTGTACGTTGATTGGGTAAAGAA 962
Qy 121 TACCGCGTGTGTGTAAGATCAAGCTTGAGAGAGGCGAGCGCGCTAAGCGCGCGCTG 180
Db 963 TATCGGCTTCTGTCAAGATTAAGTCTTGAAGAGTCAAGCCAGGCTATGCGGCTGATG 1022
Qy 181 TACACCGTGCACCTGAGAGAGCTTCAAGGTGCGGCAAGTTCGGGCAAGCTGATGATGACAGC 240
Db 1023 TACACCGTGCATCTGCTCTCTTCAAAAGTTGCTGAGAGTTGGTTCCCTTATGATGACCGT 1082
Qy 241 CTGGCGCTGCTGCGCGCGCAAGTAA 264
Db 1083 CTGGCGCTGCTGCGCGCTAAGTAA 1106

RESULT 11
US-08-349-131-2
; Sequence 2, Application US/08349131;
; Patent No. 5871974
; GENERAL INFORMATION:
; APPLICANT: HUSE, WILLIAM D.
```

```

1  TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF
2  TITLE OF INVENTION: HETEROMERIC RECEPTORS
3  NUMBER OF SEQUENCES: 75
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
6  STREET: 444 SO. FLOWER STREET, SUITE 200
7  CITY: LOS ANGELES
8  STATE: CALIFORNIA
9  COUNTRY: UNITED STATES
10 ZIP: 90072
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patent'n Release #1.0, Version #1.25
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/349,131
19 FILING DATE:
20 CLASSIFICATION: 435
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US/08/120,646
23 FILING DATE:
24 APPLICATION NUMBER: US/07/767,136
25 FILING DATE:
26 ATTORNEY/AGENT INFORMATION:
27 NAME: CAMPBELL, CATHRYN A.
28 REGISTRATION NUMBER: 31,815
29 REFERENCE/DOCKET NUMBER: P31 8882
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 619-535-9001
32 TELEFAX: 619-535-8949
33 INFORMATION FOR SEQ ID NO: 2:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 7317 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: both
38 TOPOLOGY: circular
39
40 US-08-349-131-2
41
42 Query Match: 54.5%; Score 144; DB 2; Length 7317;
43 Best Local Similarity: 71.6%; Pred. No. 1,4e-21;
44 Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
45
46 Cy 1 ATGATCAAGGTGGAGATCAAGGCCAGGCGCCAGTTCCACACCGGAGCGGCTGAGC 60
47 Db 843 ATGATTAAGTGAATTAATTAACCATCTCAAGGCCATTACTACTGCTGGTCTTCT 902
48
49 Cy 61 GCGCAGGCGCAAGCCCTACAGCCTGAAGCAGCAGCTGTACTGTGGACCTGGCAGAG 120
50 Db 903 GCGTCAAGGCAAGCCCTATTCACTGAATTAAGCAGCTTGTAGCTGATTTGGGTATATA 962
51
52 Cy 121 TACCCGCTGGTGGTGAAGATCAAGCTGACGAGGCGCAGCCGCTACGCCCCGGCTG 180
53 Db 963 TATCGGCTTCTTGTCAGATTACTCTTATGAAGGTCAAGCCCTATGCGGCTGGTGTG 1022
54
55 Cy 181 TACACCGGTGACCCCTGAGCAGCTCAAGGCGGCGCGCTGGAGGAGCTATATATGACCGC 240
56 Db 1023 TACACCGTCACTCTCTCTTTCAAGTTGTTCTAGTTCGGTTCCTTATGATTGACCGT 1082
57
58 Cy 241 CTGCGCCTGGTGCAGCCGCAAGTAA 264
59 Db 1083 CTGCGCCTGGTGCAGCCGCAAGTAA 1106
60
61 RESULT: 2
62 US-08-470-297A-2
63 Sequence 2, Application US/08470297A
64 Patent No. 6027933
65 GENERAL INFORMATION:
66 APPLICANT: HUSE, WILLIAM D.
67 TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF
68 TITLE OF INVENTION: HETEROMERIC RECEPTORS
69 NUMBER OF SEQUENCES: 76

```

```

1  ADDRESS: CAMPBELL & FLORES LLP
2  STREET: 4370 LA MOLLA VILLAGE DRIVE, SUITE 700
3  CITY: SAN DIEGO
4  STATE: CALIFORNIA
5  COUNTRY: UNITED STATES
6  ZIP: 92122
7
8  COMPUTER READABLE FORM:
9  MEDIUM TYPE: Floppy disk
10  COMPUTER: IBM PC compatible
11  OPERATING SYSTEM: PC-DOS/MS-DOS
12  SOFTWARE: Patentin Release #1.0, Version #1.25
13  CURRENT APPLICATION DATA:
14  APPLICATION NUMBER: US/38/470,297A
15  FILING DATE: June 5, 1995
16  CLASSIFICATION: 435
17  ATTORNEY/AGENT INFORMATION:
18  NAME: CAMPBELL, CATHERYN A.
19  REGISTRATION NUMBER: 31,815
20  REFERENCE/POCKET NUMBER: F-IX 1611
21  TELECOMMUNICATION INFORMATION:
22  TELEPHONE: 619-535-9001
23  TELEFAX: 619-535-9949
24  INFORMATION FOR SEQ ID NO: 2:
25  SEQUENCE CHARACTERISTICS:
26  LENGTH: 7317 base pairs
27  TYPE: nucleic acid
28  STRANDEDNESS: both
29  TOPOLOGY: circular
30
31  US-08-470-297A-2
32
33  Query Match      54.53% Score 1441 DB 3: Length 7317
34  Best Local Similarity 71.68% Pred. No. 1,46-21;
35  Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
36
37  1  ATGATCAAGGTGAGATCAAGCCGACGCGCCGATGACCAAGCCGAGGGGCTGACG 60
38  Db  843  ATGATTAAGTGAATATTAACCATCTCAAGGCGAAATTAATCTGCTTTGCGTTCT 362
39
40  61  CGGACGAGGCAAGCCCTACAGCTGACAGGAGAGTGTGCTACGTGAGTCTGGGTAAGAG 120
41  Db  903  CGTACGAGCAAGCTTTTACATGTAAGACAGATTTGTTACGTGATTTGGGTATAGAA 362
42
43  121  TACCCGCTGCTGTGAAGATCAACCTGGAGCAGGGCCAGCCCGGCTACAGCCCGCCGCTG 180
44  Db  1023  TACACCGGTGATCTGCTCTTTCATGAAGGTGGTGAAGTGGGTCAGGCTTAATGATTA 1282
45
46  161  TACACCGGTGACCGGACAGACTTCACAGTGGCCAGTGGGACCGAGATGATGACACCG 240
47  Db  963  TATCCGCTTCTTCAAGACTTACTCTTGATGTAAGGTACAGGAGCTTAAGCCCTGAGCTG 1022
48
49  241  CTGCGCTGTGTCGCGGCAAGTAA 264
50  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
51
52  265  CTGCGCTGTGTCGCGGCAAGTAA 264
53  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
54
55  265  CTGCGCTGTGTCGCGGCAAGTAA 264
56  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
57
58  265  CTGCGCTGTGTCGCGGCAAGTAA 264
59  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
60
61  265  CTGCGCTGTGTCGCGGCAAGTAA 264
62  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
63
64  265  CTGCGCTGTGTCGCGGCAAGTAA 264
65  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
66
67  265  CTGCGCTGTGTCGCGGCAAGTAA 264
68  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
69
70  265  CTGCGCTGTGTCGCGGCAAGTAA 264
71  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
72
73  265  CTGCGCTGTGTCGCGGCAAGTAA 264
74  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
75
76  265  CTGCGCTGTGTCGCGGCAAGTAA 264
77  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
78
79  265  CTGCGCTGTGTCGCGGCAAGTAA 264
80  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
81
82  265  CTGCGCTGTGTCGCGGCAAGTAA 264
83  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
84
85  265  CTGCGCTGTGTCGCGGCAAGTAA 264
86  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
87
88  265  CTGCGCTGTGTCGCGGCAAGTAA 264
89  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
90
91  265  CTGCGCTGTGTCGCGGCAAGTAA 264
92  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
93
94  265  CTGCGCTGTGTCGCGGCAAGTAA 264
95  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
96
97  265  CTGCGCTGTGTCGCGGCAAGTAA 264
98  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
99
100  265  CTGCGCTGTGTCGCGGCAAGTAA 264
101  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
102
103  265  CTGCGCTGTGTCGCGGCAAGTAA 264
104  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
105
106  265  CTGCGCTGTGTCGCGGCAAGTAA 264
107  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
108
109  265  CTGCGCTGTGTCGCGGCAAGTAA 264
110  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
111
112  265  CTGCGCTGTGTCGCGGCAAGTAA 264
113  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
114
115  265  CTGCGCTGTGTCGCGGCAAGTAA 264
116  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
117
118  265  CTGCGCTGTGTCGCGGCAAGTAA 264
119  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
120
121  265  CTGCGCTGTGTCGCGGCAAGTAA 264
122  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
123
124  265  CTGCGCTGTGTCGCGGCAAGTAA 264
125  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
126
127  265  CTGCGCTGTGTCGCGGCAAGTAA 264
128  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
129
130  265  CTGCGCTGTGTCGCGGCAAGTAA 264
131  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
132
133  265  CTGCGCTGTGTCGCGGCAAGTAA 264
134  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
135
136  265  CTGCGCTGTGTCGCGGCAAGTAA 264
137  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
138
139  265  CTGCGCTGTGTCGCGGCAAGTAA 264
140  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
141
142  265  CTGCGCTGTGTCGCGGCAAGTAA 264
143  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
144
145  265  CTGCGCTGTGTCGCGGCAAGTAA 264
146  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
147
148  265  CTGCGCTGTGTCGCGGCAAGTAA 264
149  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
150
151  265  CTGCGCTGTGTCGCGGCAAGTAA 264
152  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
153
154  265  CTGCGCTGTGTCGCGGCAAGTAA 264
155  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
156
157  265  CTGCGCTGTGTCGCGGCAAGTAA 264
158  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
159
160  265  CTGCGCTGTGTCGCGGCAAGTAA 264
161  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
162
163  265  CTGCGCTGTGTCGCGGCAAGTAA 264
164  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
165
166  265  CTGCGCTGTGTCGCGGCAAGTAA 264
167  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
168
169  265  CTGCGCTGTGTCGCGGCAAGTAA 264
170  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
171
172  265  CTGCGCTGTGTCGCGGCAAGTAA 264
173  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
174
175  265  CTGCGCTGTGTCGCGGCAAGTAA 264
176  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
177
178  265  CTGCGCTGTGTCGCGGCAAGTAA 264
179  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
180
181  265  CTGCGCTGTGTCGCGGCAAGTAA 264
182  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
183
184  265  CTGCGCTGTGTCGCGGCAAGTAA 264
185  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
186
187  265  CTGCGCTGTGTCGCGGCAAGTAA 264
188  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
189
190  265  CTGCGCTGTGTCGCGGCAAGTAA 264
191  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
192
193  265  CTGCGCTGTGTCGCGGCAAGTAA 264
194  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
195
196  265  CTGCGCTGTGTCGCGGCAAGTAA 264
197  Db  1083  CTGCGCTGTGTCGCGGCTAAGTAA 1106
198
199  265  CTGCGCTGTGTCGCGGCAAGTAA 264
2
```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/07149  
FILING DATE: 19910927  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CAMPBELL, CATHRYN A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 8882  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-535-9001  
TELEFAX: 619-535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7317 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: both  
TOPOLOGY: circular  
PCT-US91-07149-2

Query Match 54.5% Score 144 DB 5 Length 7317  
Best Local Similarity 71.6% Pred. No. 1.4e-21  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTACCAAGCCAGCCAGCTGAGC 60  
DB 843 ATGATCAAGTGGAGATCAAGCCAGCCAGCTTACCAAGCCAGCTTACCAAGCTTCT 902  
QY 61 CGCCAGGAGAGCCCTACAGCCCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCAAG 120  
DB 903 CGTCAGGAGAGCCCTATTCAGCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 962  
QY 121 TACCCGCTGTGTGAAGTCAAGCTGAGAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 180  
DB 963 TACCCGCTGTGTGAAGTCAAGCTGAGAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 1022  
QY 181 TACACGCTGACCTGAGAGAGCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 240  
DB 1023 TACACGCTGACCTGAGAGAGCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 1082  
QY 241 CTGCGCCTGTGTGCGCCAGCAAGTAA 264  
DB 1083 CTGCGCCTGTGTGCGCCAGCAAGTAA 1106

## RESULT 14

US-08-440-787A-2  
Sequence 2, Application US/08440787A  
Patent No. 5770434  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: Soluble Peptides Having Constrained,  
TITLE OF INVENTION: Secondary Conformation in Solution and Method of Making  
NUMBER OF SEQUENCES: 174  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,787A  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435

Prior Application Data:  
APPLICATION NUMBER: US 07/978,503  
FILING DATE: 12-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-IX 1586  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7320 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-08-440-787A-2

Query Match 54.5% Score 144 DB 1 Length 7320  
Best Local Similarity 71.6% Pred. No. 1.4e-21  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTACCAAGCCAGCCAGCTGAGC 60  
DB 843 ATGATCAAGTGGAGATCAAGCCAGCCAGCTTACCAAGCCAGCTTACCAAGCTTCT 902  
QY 61 CGCCAGGAGAGCCCTACAGCCCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCAAG 120  
DB 903 CGTCAGGAGAGCCCTATTCAGCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 962  
QY 121 TACCCGCTGTGTGAAGTCAAGCTGAGAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 180  
DB 963 TACCCGCTGTGTGAAGTCAAGCTGAGAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 1022  
QY 181 TACACGCTGACCTGAGAGAGCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 240  
DB 1023 TACACGCTGACCTGAGAGAGCTGAAAGAGAGAGCTGTGCTAGCTGAGCTGAGCA 1082  
QY 241 CTGCGCCTGTGTGCGCCAGCAAGTAA 264  
DB 1083 CTGCGCCTGTGTGCGCCAGCAAGTAA 1106

## RESULT 15

US-08-367 693-2  
Sequence 2, Application US/08367685  
Patent No. 6288530  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Prety, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/367,693  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,484  
FILING DATE:  
APPLICATION NUMBER: US/07/767,436  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P31 9072  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9000  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7320 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: circular  
 US-08-367-685-2

Query Match 54.5%; Score 144; DB 3; Length 7320;  
 Best Local Similarity 71.6%; Pred. No. 1,4e-21;  
 Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGGAGATCAAGCCAGCCAGCCCAAGTTCACCAACCGGACGGGCGTGAAC 60  
 DB 843 ATGATTAAGTTGAATTAACCACTCTCAAGCCCAATTACACTGCTTCGGTGTTCCT 902  
 QY 61 CCGCAGGGGCAAGCCCTACAGCTACAGGAGAGCTGTGCTACCTGGACCTGGGCAACGAG 120  
 DB 903 CGTCAGGGCAAGCCTTATTCAGTGAATGAGCAGCTTGTACGTTGATTGGGTATGAA 362  
 QY 121 TACCCCGTGGTGGAGATCAACCTGGACGAGGAGCCAGCCAGCCAGCCCGGCGCTG 180  
 DB 963 TATCGGTTCTTGCAAGATTACTCTTGAAGAGTCAGCCAGCTATGGGCTGGTCTG 1022  
 QY 181 TACACCGTGCACCTGACAGCTGACAGCTGACAGCTGAGTGGGAGCTGATGATGACCGG 240  
 DB 1023 TACACCGTTCATCTGCTCTCTTTCGAAGTGGTCAAGTGGTTCCTTATGATGACCGT 1082  
 QY 241 CTGGGCTGGTGGCCGCAAGTAA 264  
 DB 1083 CTGGGCTGGTGGCCGCAAGTAA 106

Search completed: November 7, 2003, 12:56:58  
 Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:19:21 (without alignments)

7152.412 Million cell updates/sec

Title: US-09-622-500B-3

Sequence: 264

1 atgacaaagtgagatcaaa.....gacctggcgccgcaagtaa 264

Scoring table: CLGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database:

GenBank:  
1: gb\_ba:  
2: gb\_htg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pac:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_rc:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vl:  
15: em\_ba:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_mu:  
20: em\_om:  
21: em\_ov:  
22: em\_ov:  
23: em\_pac:  
24: em\_ph:  
25: em\_pl:  
26: em\_rc:  
27: em\_sts:  
28: em\_un:  
29: em\_vl:  
30: em\_vl:  
31: em\_htg\_hum:  
32: em\_htg\_iny:  
33: em\_htg\_other:  
34: em\_htg\_pln:  
35: em\_htg\_rnd:  
36: em\_htg\_rnd:  
37: em\_htg\_vrt:  
38: em\_sy:  
39: em\_htgo\_hum:  
40: em\_htgo\_mus:  
41: em\_htgo\_other:

Pred. No. is the number of results predicted by chance to have a

Score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	FR	ID	Description
1	21	9.0	230905	2	AC132180	AC132180 Rattus no
2	20	7.6	674	9	HSB136570	AC132180 Homo sapi
3	20	7.6	15124	2	AC108763	AC108763 Oryza sat
4	20	7.6	347950	1	AF003013	AF003013 Mesocric
5	19	7.2	3624	3	AF132169	AF132169 Drosophi
6	19	7.2	3773	1	PEENORC	PEENORC Drosophi
7	13	7.2	10711	1	AE004489	AE004489 Pseudom
8	13	7.2	10024	1	AE012081	AE012081 Xanthom
9	13	7.2	11854	1	AE012540	AE012540 Xanthom
10	19	7.2	14616	9	HS064621	HS064621 Human DNA
11	19	7.2	20686	2	AC017508	AC017508 Drosophi
12	19	7.2	74433	2	AC015098	AC015098 Drosophi
13	19	7.2	107056	9	AL596822	AL596822 Human DNA
14	13	7.2	135055	2	AC010556	AC010556 Drosophi
15	13	7.2	135055	2	AC011909	AC011909 Drosophi
16	19	7.2	144305	2	AC015056	AC015056 Mus muscu
17	19	7.2	150959	2	AC019250	AC019250 Homo sapi
18	19	7.2	164617	2	AC102624	AC102624 Mus muscu
19	19	7.2	166195	10	AL928797	AL928797 Mouse DNA
20	19	7.2	168514	3	AC023350	AC023350 Drosophi
21	19	7.2	174920	3	AC017501	AC017501 Drosophi
22	19	7.2	177205	2	AC075217	AC075217 Mus muscu
23	19	7.2	177577	2	AC022349	AC022349 Drosophi
24	19	7.2	188225	2	AC108587	AC108587 Rattus no
25	19	7.2	189461	2	AC118247	AC118247 Mus muscu
26	19	7.2	190598	10	AC038892	AC038892 Mus muscu
27	19	7.2	202405	2	AC124556	AC124556 Mus muscu
28	19	7.2	203261	2	AC011374	AC011374 Homo sapi
29	19	7.2	225031	2	AC131174	AC131174 Mus muscu
30	19	7.2	225671	2	AC132397	AC132397 Mus muscu
31	19	7.2	226273	2	AC111408	AC111408 Rattus no
32	19	7.2	235936	2	AC036473	AC036473 Rattus no
33	19	7.2	242510	2	AC037243	AC037243 Rattus no
34	19	7.2	255792	2	AC138101	AC138101 Rattus no
35	19	7.2	261977	2	AC037564	AC037564 Rattus no
36	19	7.2	266511	2	AC133768	AC133768 Rattus no
37	19	7.2	273980	2	AC111446	AC111446 Rattus no
38	19	7.2	277331	2	AC110151	AC110151 Rattus no
39	19	7.2	294272	3	AE033595	AE033595 Drosophi
40	19	7.2	298700	1	AP035951	AP035951 Bradyniz
41	19	7.2	305469	3	AE030501	AE030501 Drosophi
42	19	7.2	305550	3	AP030503	AP030503 Streptomy
43	18	6.8	466	8	AY063045	AY063045 Arabidops
44	18	6.8	587	6	BD155522	BD155522 Primer fo
45	18	6.8	646	8	AY034898	AY034898 Arabidops

# ALIGNMENTS

RESULT:  
AC132180  
LOCUS  
DEFINITION Rattus norvegicus clone CH23C.123H15, WORKING DRAFT SEQUENCE, 3  
AC132180  
ACCESSION  
VERSION AC132180.3 G1:23664858  
KEYWORDS HTG, HTGS, PHASE1, HTGS, DRAFT, HTGS, FULLTOP.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Buxaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,  
Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae,  
Rattus.  
REFERENCE 1 (bases 1 to 230905)  
Muzny,D,Marie, Metker,M, Lee, Abramson,S., Adams,C., Alder,J.,





```

JOURNAL      of the human genome
MEDLINE      Nucleic Acids Res. 30 (14), 3163-3170 (2002);
PUBMED       1213767
REFERENCE     2 (bases 1 to 674)
AUTHORS      Zabarovsky, B.R.
TITLE        Direct Submission
JOURNAL      Karolinska Institute, Theorells vag, 3, Box 280, Stockholm S-171 77,
Sweden
FEATURES     Location/Qualifiers
             1..674
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /clone="NR1-P1248"
BASE COUNT   125 a 204 c 226 g 119 t
ORIGIN
Query Match  7.6%; Score 20; DB 9; Length 674;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 162 CGCTACGCCCCCGGCGCTGT 181
      |||||
Db 50 CGCTACGCCCCCGGCGCTGT 31

RESULT 3
AC108763      155124 bp DNA linear HTG 31-JAN-2002
LOCUS         Oryza sativa (japonica cultivar-group; chromosome 9 clone
DEFINITION    OSJNB0004A05, *** SEQUENCING IN PROGRESS ***, 15 ordered pieces.
VERSION       AC108763
KEYWORDS      HTG; HTGS; HTGS43968
SOURCE        Oryza sativa (japonica cultivar-group)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Cryzeae; Oryza.
REFERENCE     Yoon, J.-H., Hahn, J.-H., Yun, D.-W., Lee, J.-S., Lee, M.-C., Eun, N.-Y.
AUTHORS       and Kim, H.-I.
TITLE         Oryza sativa BAC OSJNB0004A05 genomic sequence
JOURNAL       Unpublished
REFERENCE     Hahn, J.-H. and Kim, H.-I.
AUTHORS       Direct Submission
TITLE         Submitted (31-JAN-2002) Rice Genome Sequencing Project, National
JOURNAL       Institute of Agricultural Science and Technology (NIAS), RDA, 249
               Seodun-dong, Suwon 441-707, Korea (E-mail: jhahn@da.go.kr,
               Tel: 82-31-290-0309, Fax: 82-31-290-0308)
COMMENT       * NOTE: This is a 'working draft' sequence. It currently
               * consists of 15 contigs. Gaps between the contigs
               * are represented as runs of N. The order of the pieces
               * is believed to be correct as given, however the sizes
               * of the gaps between them are based on estimates that have
               * been provided by the submitter.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.
               * 1 7850: contig of 7850 bp in length
               * 7851 7950: gap of unknown length
               * 7951 17664: contig of 9714 bp in length
               * 17665 17665: gap of unknown length
               * 17665 39434: contig of 21670 bp in length
               * 39435 39435: gap of unknown length
               * 39535 44003: contig of 4469 bp in length
               * 44004 44103: gap of unknown length
               * 44104 55602: contig of 11499 bp in length
               * 55603 55702: gap of unknown length
               * 55703 107219: contig of 51517 bp in length

```

```

* 107220      107319: gap of unknown length
* 107320      114663: contig of 7344 bp in length
* 114664      114763: gap of unknown length
* 114764      115098: contig of 2335 bp in length
* 117099      117198: gap of unknown length
* 117199      131565: contig of 14767 bp in length
* 119666      132065: gap of unknown length
* 120066      134114: contig of 2249 bp in length
* 134115      134315: gap of unknown length
* 134315      136027: contig of 4593 bp in length
* 136028      139137: gap of unknown length
* 139138      149593: contig of 10492 bp in length
* 149600      151781: contig of 2082 bp in length
* 151782      151881: gap of unknown length
* 151882      153161: contig of 1280 bp in length
* 153162      153261: gap of unknown length
* 153262      155124: contig of 1963 bp in length.
             Location/Qualifiers
             1..155124
             /organism="Oryza sativa (japonica cultivar-group)"
             /mol_type="genomic DNA"
             /variety="Nipponbare"
             /db_xref="taxon:39947"
             /chromosome="9"
             /clone="OSJNB0004A05"
BASE COUNT   43170 a 33383 c 33969 g 43152 t 1450 others
ORIGIN

Query Match  7.6%; Score 20; DB 2; Length 155124;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 194 TGAGCAGCTTCACGCTCGGC 213
      |||||
Db 81541 TGAGCAGCTTCACGCTCGGC 81560

RESULT 4
AC103013      347350 bp DNA linear BCT 15-MAY-2002
LOCUS         Mesorhizobium loti DNA, complete genome, section 20/21.
DEFINITION    AP003013 SA000012
ACCESSION     AP003013.2 GI:14027324
VERSION       AP003013.2 GI:14027324
KEYWORDS
SOURCE        Mesorhizobium loti
ORGANISM      Mesorhizobium loti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE     Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
AUTHORS       Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
               Kishida, Y., Kiyokawa, C., Kohara, Y., Matsumoto, Y., Matsuno, A.,
               Yoshizuki, Y., Nakayama, S., Nakazaki, N., Saito, S., Sugimoto, M.,
               Takeuchi, C., Yada, M. and Tabata, S.
TITLE         Complete genome structure of the nitrogen-fixing symbiotic
               bacterium Mesorhizobium loti
JOURNAL       DNA Res. 7 (6), 331-338 (2000);
MEDLINE       21082910
PUBMED        11214968
REFERENCE     Kaneko, T.
AUTHORS       Direct Submission
TITLE         Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
JOURNAL       Institute, The First Laboratory for Plant Gene Research, Yara
               1532-3, Kasarazu, Chiba 292-0812, Japan
               (E-mail: kaneko@kazusa.or.jp,
               URL: http://www.kazusa.or.jp/rhizobase/,
               Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934).
COMMENT       On May 11, 2001 this sequence version replaced gi:11994388.
             Location/Qualifiers
             1..347350
             /organism="Mesorhizobium loti"

```

```

/mol_type="genomic DNA"
/strain="MAFF30399"
/db_xref="taxon:381"
complement(78..1037)
/gene="mll7923"
complement(78..1037)
/gene="mll7923"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB3594.1"
/db_xref="GI:14027325"
/transl_table="MELISQVNLNSLKIVESAASHGNTFRAGSECFITASAVSCWKS
:EQCLAFKIQRGNAVSLTETSETVSVREALERIVAASWEATQSGEHVXKGLVL
FIFAAAWFPAJPEFQRRYCDIIVMRVSTSYATHEFTSTQFDEIRYGGSHFNGLPSDL
LFREDTTFVSRKLFHEVAGDKPVSVKPEDDLHFTLLHSDTCTQNMQMLGFAGASF
VLSRTKSYPFDSOMGYEANAAGMGFAVANKRAYMADIRAEKLVAFFAVHPHSAGWY
FVSPHKCLAAKVKLLFKQWVAEAAJTORQDSEIRDLATEAY"
1150..1986
/gene="mll7924"
1150..1986
/gene="mll7924"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3595.1"
/db_xref="GI:14027326"
/transl_table="XNAHTTPELRYAVSREAI:GHETANKVSVFGVAQYLHGVDPALJ
AAIEFAAKIKLDSHAYKHKJDLFTITGAPRFPEIKELLHDKLRJERLDMMGTKLEP
YPLISVGSVTFPMKQVAGWECGVPTTELPLAISNPLVGGHLEIVCDSEVSGRA
LDSGSEIPNNRMDHKMNVATLGQFVJHRTAPIQVGERVPTVLNORVAKFYV
DQNMFMFLAANDNDHREWNELAEVDKWCNLPAYRFEABHTPPAFASSETQVSGNARG
SK"
2032..2958
/gene="mll7925"
2032..2958
/gene="mll7925"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3596.1"
/db_xref="GI:14027327"
/transl_table="MVQHRDRAGRILFRHGSSISALVFAAGAVALKSTNA:VGKSLI
ATPVSQVQPLQAGALVFLALIPMSSEAPSPVAPAAVAALAVGIGLWGTMVLCY
IAPRSVIANLVAWYFPLMVAAYAFDNPAPRALGLAAV:GFVGVALLVMSGRE
NMFQGLVGIFAFAGALCMATYSVMVGRLATSPRELLDPSSEVGVVALALLWSAGEG
VAMPAGMDLALGVLGAGPMGLGYFWSRAKLEGGKVAWVAWYLTIP:ASTGLLTUSG
EQLTTTAAGAVLVIGSIAVGLERSAQDYV"
2951..3154
/gene="mll7926"
2951..3154
/gene="mll7926"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3597.1"
/db_xref="GI:14027328"
/transl_table="MPCDNERLARQEAHMLIKFGEAPAIYAAKMAEKAIECKDFRSC
ARWRRIELADAKRITKSAASK?"
3307..3759
/gene="mll7927"
3307..3759
/gene="mll7927"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB3598.1"
/db_xref="GI:14027329"
/transl_table="XKLSNTEQXTV:ELTADIVAVYGNPLPASGLPDLIASVSAS
VKLAGAVVESPL:VPAVNPKSVFPEDYICLEDGKKFKSLXRHLATCYG:GFDYR
AKWGLPDYFVVAWYNATRSALAKSTGLGRKKPAAAPAAVAKKAKA"
3916..4065
/gene="mll7929"
3916..4065
/gene="mll7929"

```

---

```

CDS
3916..4065
/gene="msr7929"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3599.1"
/db_xref="GI:14027330"
/transl_table="MNCNAYFEAKGRSALVLQXVFTLFLELLGGSKPLDRNQHCRH
SLRPR"
complement(4476..4730)
/gene="msl7931"
complement(4476..4730)
/gene="msl7931"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3600.1"
/db_xref="GI:14027331"
/transl_table="MRSERSLMISYAALLHRTSTLRKALPASWPDVRVKTGGSGSALRH
QPGGIKWRPQDFCCVADRKTAGTDIEDTCLRNLYFYSC"
4699..4699
/gene="msl7932"
4699..4699
/gene="msl7932"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding component"
/protein_id="BAB3601.1"
/db_xref="GI:14027332"
/transl_table="MLIINDLSRYAGRLDDCHASLTDFPAGTKAGLVGRNQTGTLTF
KATGDFSESTGSLRKTIRIGOVAGBAPTETPLIEIIVLKAQVETALIEBKAT
DPRHAIQHMELAD:DAHSABSBAT:LAG:GFDDAAGRPASSFGMRMRVALAAV
LSEPDLLIDEPFNYLDLEGTJWLENYVSKYPTHVLL:SHDRLLLRANVSIHVLDQ
KCLTFRGGYDQFEQYTEQKEQKGRVKOEARKHSEFVERFRAKAKARQQR
IKALEMKPIAALVNDSPFSEFPYVASTPVALNVMNVGYTEGQPLKMTLRI
DADRIALLGANGKSTFAKLLSGRIKQESVTVPGLKVAIFAQHLDLRPEEN
AYEHVRRMLPAPSKSVGRVAGFGLTEKMTFAKOLSGEKARL:MGLSAFSGPNJ
FILDEPTNHLIDQRESLIHALNEFPAGVILI:SHDRHLLLEATRLWVKDGAHPVE
GDLECYKTLVTGVSGRGRKREAKKASKEPRDAAARAAFEFLAKEIRATEALMDR
IKRR:DGIEDELANFAIYEKDPSTATLAKERSQLAQTJAGHEEKWLTMSABYSEGTA
E"
6667..8178
/gene="mll7933"
6667..8178
/gene="mll7933"
/notes="similar to formate dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB3602.1"
/db_xref="GI:14027333"
/transl_table="MNOHKLRI:GHSACPHDPCSTCALEVELLDNRI:GRVHGAKANT
YTSGVVCAKVARVADRHHPRDLKPLIRAGAKGQVQWQBSAAALDLVAEKIAAE
EKYQSETVMFYFAASC:AAVTGSQVGGGAPHNSGIFKQNEVLGSTRMROPDIR
YLDHNR:GPV:GASDALYGGPPVTAML:QNTNPNVAPQRLVQKQELRDLQFACVH
EQMTD:ARLADVLVLPATYFLEHDVVKGNGQHT:IGPKLI:EPPEGTHTFVIEQL
ARLQVADRPGGLTQOOHDIILKRGLSGFSGLKEQKWYDLOQFETAHTFKGFGH
SDGPRPADWTGGAAPNRPKSVGLFGPVRLLPEFPDHVZLIEVADDAHPRP:JTSP
ARNFLNST:FASTVSKSEKGPALLLHPDDADLGLADGDRVEYGNQPGDLVLAKEFF
DQ:KAGVVASG:WPNSSHERGGINVLITGACAFYGGAAFHONKVVLRKAG"
complement(8203..9390)
/gene="mll7934"
complement(8203..9390)
/gene="mll7934"
/notes="similar to dimethylamine monooxygenase"
/codon_start=1
/transl_table=11
/protein_id="BAB3603.1"
/db_xref="GI:14027334"
/transl_table="WFLTFSGRE:SRBARLDTTTVALIGAGPAGLAAAC:RCAGVD
PIL:EKESQAPAMPRHVRVHLHTTKYSSLPFPVFPKPYRYVFRALFYV:LDAYA
GRFLRPOFGETVKAATQCGMGRVDAASGFLRAKTVIAGSYNAEPURPAPAGITP
TGKTLHSADYRNKAF:FAQGSVLV:MGNT:GAELALDAENGAQFTISVREGVHVPRE
LPGVFIQVGMMAARLGPQRNDALFFVILDLVLGRDLKYG:KRPPQGG:LEQIVASR:

```

PVIDVGTGKIREGAIKVAPDITEISQGRFADGKHGEFDEAIFATGYAPGYARFLE  
PGIQDRSGVTPKASDLGLYLIGHFNNAVTLLEIGIEQAIAADDIHRNLNRKAAEI  
LPV"

gene complement (9426, .9905)

/gene="ml17935"

CDS complement (9426, .9905)

/gene="ml17935"

/codon\_start=1

/transl\_table=1

/product="transcriptional regulator"

/protein\_id="BAB3f84.1"

/db\_xref="GI:14027335"

/translation="VPESSPPDLSFDLAILVQLQONTTPQRLIGAVNLAPAVQRR

KRMQAGV:ASNAV:EPAAVGPKITTFVEVELESERTLIDAARQROSTPEVQCC

YVVTGEADFLIVFVADMGAVEALTRKLFSSNVKRTFFVADRVKVGSLTVPD"

10133, .11218

gene /gene="ml17936"

CDS 10133, .11218

/gene="ml17936"

/note="hypothetical protein"

/codon\_start=1

Query Match 7.6%; Score 20; DB 1; Length 347950;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 20; Conservative C; Mismatches 0; Indels 0; Gaps C;

QY 201 CTTCAAGGTGGCCAGTTC 220

||||| ||||| ||||| |||||

Db 21:819 CTTCAAGGTGGCCAGTTC 211838

RESULT 5

AF132169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

mol\_type="mRNA"

/db\_xref="taxon:722"

/cdate="1021405"

454, .3093

/codon\_start=

/product="unknown"

/protein\_id="AAD34757.1"

Location/Qualifiers

1..3624

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:722"

/cdate="1021405"

454, .3093

/codon\_start=

/product="unknown"

/protein\_id="AAD34757.1"

Location/Qualifiers

1..3624

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:722"

/cdate="1021405"

/db\_xref="GI:4972724"

/db\_xref="FLYBASE:FBcd0021405"

/translation="MAINQGYKYSICAIIDEVDERSFKLHNSQDBIGKQKQLOQIEGS

ETAPUTITQCKJHMDGSHLTPKDLAAQDAALNEAAGGAGGRTPAAITPNSHPSA

NGLRALLPGRSFDALMSLRKRILYVTTACICALLLLIILLAPWEPVFLRA

PLCDKEGVSSRQLLWANSQSKCHETYEACNFASDYANHEFYV:KRGEMNYET

YNEVOELNELNRFIAE:PNISQSGSVESVSSLYRSCTDVLKQSDLLKLRANAV

YMQOAFRDSNBLQSWERYKVLVLCAGHGIPEYAVTVVENGESKPHDVITLDEGELG

/PDKYFYGNDADEEVVRYKLLLDLDFAINMGIVSRADLFADDFIHYERRIVNH:EDA

KADANRPINKLRLADLKIKAUPSILLESJQAIFPKTKITEDEVLVKDPVEMHALSV

LLTSDSKPINKFVMSLRKYLPHLSKEYRILAEITFCHALYGRFASYPRMLICSKV

RDWJFAVDA:QQPPRQCFETM:SKSRGLNKTQNEELLKLMFSLRNQKDS:EQA

KMLSPAAQF:QCKLNEMLQFG:PDEVLCQPTY:AOYVNELI:LNKVFVEHLE:WT

FRRSMEKLGPLA:LQVIVSEMVRDNPQAIAYSNKLM:LLVSKVLSSKAVVDYVP

VAWFAELGTDL:ELALDNSTFELLQNTOSTDI:SEAAPE:RYACPVNCC:AAQGPFR

LYHELNEUSSNALSKFHTLSAARTAAALSNFVNAIDAGSAIQSGIGDQATYEALG

LTRRLRFGSRFENELTUSYMQOHCSTTIADQYARIKPLVESQIAESYLEFNATV

CHIOFLPSTNCAAAEATCSNLL"

BASE COUNT 995 a 911 c 884 g 834 t

ORIGIN

Query Match 7.2%; Score 19; DB 3; Length 3624;

Best Local Similarity 100.0%; Pred. No. 136-02;

Matches 19; Conservative C; Mismatches 0; Indels 0; Gaps C;

QY 138 GATACCCCTGGACGAGGCG 156

||||| ||||| ||||| |||||

Db 1428 GATACCCCTGGACGAGGCG 1446

RESULT 4

AF132169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

/db\_xref="GI:4972724"

/db\_xref="FLYBASE:FBcd0021405"

/translation="MAINQGYKYSICAIIDEVDERSFKLHNSQDBIGKQKQLOQIEGS

ETAPUTITQCKJHMDGSHLTPKDLAAQDAALNEAAGGAGGRTPAAITPNSHPSA

NGLRALLPGRSFDALMSLRKRILYVTTACICALLLLIILLAPWEPVFLRA

PLCDKEGVSSRQLLWANSQSKCHETYEACNFASDYANHEFYV:KRGEMNYET

YNEVOELNELNRFIAE:PNISQSGSVESVSSLYRSCTDVLKQSDLLKLRANAV

YMQOAFRDSNBLQSWERYKVLVLCAGHGIPEYAVTVVENGESKPHDVITLDEGELG

/PDKYFYGNDADEEVVRYKLLLDLDFAINMGIVSRADLFADDFIHYERRIVNH:EDA

KADANRPINKLRLADLKIKAUPSILLESJQAIFPKTKITEDEVLVKDPVEMHALSV

LLTSDSKPINKFVMSLRKYLPHLSKEYRILAEITFCHALYGRFASYPRMLICSKV

RDWJFAVDA:QQPPRQCFETM:SKSRGLNKTQNEELLKLMFSLRNQKDS:EQA

KMLSPAAQF:QCKLNEMLQFG:PDEVLCQPTY:AOYVNELI:LNKVFVEHLE:WT

FRRSMEKLGPLA:LQVIVSEMVRDNPQAIAYSNKLM:LLVSKVLSSKAVVDYVP

VAWFAELGTDL:ELALDNSTFELLQNTOSTDI:SEAAPE:RYACPVNCC:AAQGPFR

LYHELNEUSSNALSKFHTLSAARTAAALSNFVNAIDAGSAIQSGIGDQATYEALG

LTRRLRFGSRFENELTUSYMQOHCSTTIADQYARIKPLVESQIAESYLEFNATV

CHIOFLPSTNCAAAEATCSNLL"

BASE COUNT 995 a 911 c 884 g 834 t

ORIGIN

Query Match 7.2%; Score 19; DB 3; Length 3624;

Best Local Similarity 100.0%; Pred. No. 136-02;

Matches 19; Conservative C; Mismatches 0; Indels 0; Gaps C;

QY 138 GATACCCCTGGACGAGGCG 156

||||| ||||| ||||| |||||

Db 1428 GATACCCCTGGACGAGGCG 1446

RESULT 4

AF132169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

PVIDVGTGKIREGAIKVAPDITEISQGRFADGKHGEFDEAIFATGYAPGYARFLE

PGIQDRSGVTPKASDLGLYLIGHFNNAVTLLEIGIEQAIAADDIHRNLNRKAAEI

LPV"

complement (9426, .9905)

/gene="ml17935"

complement (9426, .9905)

/gene="ml17935"

/codon\_start=1

/transl\_table=1

/product="transcriptional regulator"

/protein\_id="BAB3f84.1"

/db\_xref="GI:14027335"

/translation="VPESSPPDLSFDLAILVQLQONTTPQRLIGAVNLAPAVQRR

KRMQAGV:ASNAV:EPAAVGPKITTFVEVELESERTLIDAARQROSTPEVQCC

YVVTGEADFLIVFVADMGAVEALTRKLFSSNVKRTFFVADRVKVGSLTVPD"

10133, .11218

gene /gene="ml17936"

CDS 10133, .11218

/gene="ml17936"

/note="hypothetical protein"

/codon\_start=1

Query Match 7.6%; Score 20; DB 1; Length 347950;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 20; Conservative C; Mismatches 0; Indels 0; Gaps C;

QY 201 CTTCAAGGTGGCCAGTTC 220

||||| ||||| ||||| |||||

Db 21:819 CTTCAAGGTGGCCAGTTC 211838

RESULT 5

AF132169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

mol\_type="mRNA"

/db\_xref="taxon:722"

/cdate="1021405"

454, .3093

/codon\_start=

/product="unknown"

/protein\_id="AAD34757.1"

Location/Qualifiers

1..3624

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:722"

/cdate="1021405"

454, .3093

/codon\_start=

/product="unknown"

/protein\_id="AAD34757.1"

Location/Qualifiers

1..3624

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:722"

PVIDVGTGKIREGAIKVAPDITEISQGRFADGKHGEFDEAIFATGY



	FEATURES	location/Qualifiers
	source	1..10711 /organism="Pseudomonas aeruginosa PAO1" /mol_type="Genomic DNA" /strain="PAO1"
gene		/db_xref="taxon:28984" 139..921 /gene="nirQ" /locus_tag="PA0526"
CDS		139..921 /gene="nirQ" /locus_tag="PA0526"
		/note="protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Cytoplasmic; Subcellular localization confidence: Class 1" /codon_start=1 /transl_table=1 /product="regulatory protein NirQ" /protein_id="AAO3909.1" /db_xref="GI:9946386"
gene		/translation="NRDATPEVATGHEIEVFEEAWRHGVLLKGPCTGGKTFNVCY NARRELDPYSVACHEDLCADLGRHLHICAGDWQDGPETRAVRGGCYDEUVE ARODTIVAHPADORRELYERTGEITLCAPPSMLVSNYPGNQLIKELSPHQF FVALREDPAQOEIRILVSGGAETLAQRVGCGALRRLEOHDEEVASTRLLIF AARLICGDVDREACRVALAEPLSDCEATVAALMDIVLHA"
CDS		956..1483 /locus_tag="PA0521" /note="synonym: nirQ" 956..1483 /locus_tag="PA0521"
		/note="protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif structural feature or limited sequence similarity to an experimentally studied gene)" /codon_start=1 /transl_table=1 /product="probable cytochrome c oxidase subunit" /protein_id="AAO3910.1" /db_xref="GI:9946387"
gene		/translation="WKFILAECSVSLLLAFAGAQAQWQPEFRAGREQJDRWSGL GVALSSGMIAAACHAVHVRPRQASGWLSAALLALVYGVLKHEYGRLDGLD IHDFTFYMLGTGFHFRLV.LGLSLGWAQACWRRCPTPCACSGLSSVLYWHMV D.VVYLJFLPVVVLQ"
CDS		2491..2748 /locus_tag="PA0522" /note="synonym: nirP" 2491..2748 /locus_tag="PA0522"
		/note="protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAO3911.1" /db_xref="GI:9946388"
gene		/translation="WSTLTCLWALLAANTCVLLGGAGDSPWLLANVNCAYAKGWJ GERMEWNAHAPALMRLLAWPFLMAVGAALYLRNN"
CDS		1838..2278 /gene="nirC" /locus_tag="PA0523" 1838..2278 /gene="nirC" /locus_tag="PA0523"
		/note="protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane-associated; Subcellular localization confidence: Class 1; Alt protein name: cytochrome c subunit of nitric oxide reductase" /codon_start=1 /transl_table=1 /product="nitric-oxide reductase subunit C" /protein_id="AAO3912.1"
gene		/db_xref="GI:9946389" /translation="YSTFTKGARNYFGGSVEFFLLFLALTYHTFKTLPSFTNEAQ MSAAYWGKUWEONNVGCHTLLGGAFAEPGLGNVVRGSGEGENTFOAXXNTIQ PLRVGPRAPQPCHLSEQVEDLAEFLKWSSKIDTNCPNPKEG"
CDS		2278..3678 /gene="nirB" /locus_tag="PA0524" 2278..3678 /gene="nirB" /locus_tag="PA0524"
		/note="protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 1; Alt protein name: cytochrome B subunit of nitric oxide reductase" /codon_start=1 /transl_table=1 /product="nitric-oxide reductase subunit B" /protein_id="AAO3913.1" /db_xref="GI:9946390"
gene		/translation="MNSPNSLKFAQAVAKPYEFALIFVCOILFGLIMGLQVVYVG DPJFAPFNVARMHVTNLVWLDFMGAAIYVPESDCSYSPKLAWLFVFWFA AAGVTLTGLLVYPAGLARLTGNEWPMGREFEQPTISKAGIVIALGFLPNVM TVJRGRTYA.SMWLYTGLIGLALFPFSFYNPENLRDKFYWWVHLWAVEGYMLIM GAIALPYLVITGVDRVIEKKLVII.AVALISGITGHVFMICGVPGV.MLGSVF SALEPLVPEAMV.PAENTINRRERDYENRAVALAMGTTMAFLANGNGEMHTLAP VNYTHOTQLTAAGHMAYGAYAXVMIIISYMPRLKIGIEMANNRQOVLEMWGFV LMTVAWFILFUSAGVQLQRPADGAATPWATODQAIFYWREGAGGVFLI GLVAYLS.FRRGKAAN"
CDS		3680..5518 /locus_tag="PA0525" /note="synonym: nirD" 3680..5518 /locus_tag="PA0525"
		/note="protein name confidence: Class ? (function proposed based on presence of conserved amino acid motif structural feature or limited sequence similarity to an experimentally studied gene)" /codon_start=1 /transl_table=1 /product="probable dinitrification protein NirD" /protein_id="AAO3914.1" /db_xref="GI:9946391"
gene		/translation="MAPAVELEVGAHKHRFITRHASGEFEARVLTLSMRPP.CMI FRALGGAPYALEATFARRLLRTW.QRWAGTCQAPVSWNGDSLRLPDESJAVTPQ AEJARELYRWALAASAGLPURMAQDNQARQQLLDAYPAUPRVARLYAAHLRQPF SLDDCPAADAEIALRAEAPEGVERFERVAMPVPLVYPERWTGPSASAEDG EEAAGAKQEVARSARGAKKEARIERNSENRLLFRLSLAWSHELALRCSDDG DPGARVAEDLDVLSRSORTKGG.RLDDLPADAYDOLPUGGLKLPEWDYRQCR LLSPHVLQPMRPGATSLSLPAHEKTALHLPQFACRDGQRRLRQOQODEILDV AWLDFQERRRGSTQGPLRPRDLACLJLADLMSFAYLDDORRVIDSILO SLLFLGALCALGDPFALYCFSSVRQCQWKQFDGEGYGEVGRVLAUSPGYT RMGAATRRASQGGQPKRLLLLSDDGRNDJDRIEGRYDIETQRAVIEARSQGL VPFCITIDKEADYLPYIPGAGFALVRAGOLPERLLQYRRLRR"
CDS		complement(5525..5719) /locus_tag="PA0526" complement(5525..5719) /locus_tag="PA0526"
		/note="protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="hypothetical protein"
	Query Match	7.28; Score 19; DS 1; length 10711;
	Best Local Similarity	100.0%; Pred. No. 1,3e+02;
	Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	144	GCTACTGCTGAAGATCAC 144
Ox	2943	GCTGCTGCTGAAGATCAC 2943

## RESULT 8

## AE012381

## LOCUS

## DEFINITION

## Xanthomonas axonopodis pv. citri str. 306

## the complete genome.

## ACCESSION

## AE012381

## VERSION

## AE012381.1

## KEYWORDS

## SOURCE

## Xanthomonas

## ORGANISM

## Xanthomonas axonopodis pv. citri str. 306

## Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

## Xanthomonadaceae; Xanthomonas.

## REFERENCE

## 1 (bases 1 to 11004)

## da Silva, A.C.R., Ferro, J.A., Peirach, F.C., Farah, C.S., Furlan, L.R.,

## Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

## Cr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

## Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,

## Chambergo, P., Clapina, L.P., Cicarelli, R.M.B., Coutinho, L.,

## Ferreira, R.C.C., Ferro, M.F., Formighieri, E.F., Franco, X.C.,

## Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

## Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A.,

## Madelra, A.X.B.N., Martinez-Rossi, N.M., Martins, E.C., Medeiros, J.,

## Menck, C.F.M., Miyaki, C.Y., Xoor, D.H., Moreira, L.M., Novo, M.T.M.,

## Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,

## Ressi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,

## Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos

## Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

## Kitajima, J.P.

## TITLE

## Comparison of the genomes of two Xanthomonas pathogens with

## differing host specificities

## JOURNAL

## Nature 417 (6897), 459-463 (2002)

## PUBMED

## 22022145

## 22022147

## 2 (bases 1 to 11004)

## da Silva, A.C.R., Ferro, J.A., Peirach, F.C., Farah, C.S., Furlan, L.R.,

## Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

## Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

## Camargo, J.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,

## Chambergo, P., Clapina, L.P., Cicarelli, R.M.B., Coutinho, L.J.,

## Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,

## Ferreira, R.C.C., Ferro, X.C., Formighieri, E.F., Franco, M.C.,

## Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

## Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A.,

## Madelra, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Medeiros, J.,

## Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

## Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,

## Ressi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,

## Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos

## Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

## Kitajima, J.P.

## Direct Submission

## Submitted (28-NOV-2001): Departamento de Bioquímica, Universidade de

## Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,

## Brazil

## Location/Qualifiers

## 1..11004

## /organism="Xanthomonas axonopodis pv. citri str. 306"

## /mol\_type="genomic DNA"

## /strain="306"

## /db\_xref="taxon:90486"

## /note="pathovar: Citri"

## 211..679

## /gene="XAC4278"

## 211..879

## /gene="XAC4279"

## /note="identified by sequence similarity; putative; ORF

## located using Blastx/Glimmer/Genemark"

## /codon\_start=1

## /transl\_table=1

## /product="conserved hypothetical protein"

```

/protein_id="AA03113.1"
/db_xref="GI:2110719"
/translation="MTE-SPPVJTTAARVVGGLIEKFAATTDPAYPTVVAQVAAN
QTARPEVNLQGTGVVHARLQETLGLVPPQFSRRAERYEHLGSLVPLTRQVAL:
GILLRGACTLQELVARLERLARFENDADMPHLDL:ORGLAOLPRASGOREDRVA
HL:SGELCVLA:QNAARARARSRSSTSDCLEARMOALEATVVE:QELAAVQARLD
AAGA"
complement(1156..2271)
/gene="XAC4279"
complement(1156..2271)
/gene="XAC4279"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA03114.1"
/db_xref="GI:2110720"
/translation="NHRILWSTORITARTNRVVGSPKDACMSLVVNAVSTLRTPLWP
DLPQAVQHRDHADPALVDH:HGFGVYVNGAGDSQMTQSEVH:MRHVQVREHESFC
VDAGFGLACWASQANAVCLADSVDRPGRVLISGEPALDGAQVPPFDALQR
RAQKSLTAQGIKRVPPSPVPVGEARVDAVVGSM:ALFAVALURAEILAGDT
PPSLDEVTR:PGVAALSPQERAFQAQAPSAQALNFGWYELAVLQWALGADT
LPSP:ALCEVPLVQAQALNQAARPTVVARPPLPELLMDLRHLRLH:WAVRQAGSA
CPVAGIVGVYVYRHYALNLLHFDAAEWDSVDTP"
complement(12527..3081)
/gene="XAC4280"
complement(12527..3081)
/gene="XAC4280"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="acetyltransferase"
/protein_id="AA03115.1"
/db_xref="GI:2110721"
/translation="MMPSCAVTV:QTPRLRLTLPDRDADAMLELVNDPQFIAGIN
DRG:ITRQCAHLCWQAQAHVGFARHAWETREDAA:GTLLGLLRETLPVPHLG
FALLPAYKGVVTEAGRAVLIDYAPVLGLSOLCAIVSPDCAESIRAEIKLQHQHGL
RV:SPETAFVYVADLGFVGRAG"
complement(13159..15135)
/gene="rep"
/note="synonym: XAC4281"
complement(13159..15135)
/gene="rep"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="Arp-dependent DNA helicase"
/protein_id="AA03116.1"
/db_xref="GI:2110722"
/translation="NYGLNPSQAAVHCRGRLVLVLAGSGSKTRVIEKIAHLIAG
RYPAP:AAITFTNKSAREVRVAK:RGDGAELTICTHAGLKFTQ:EHAAAGL
KRGFS:FSDDAAQ:KOLMHGAKPDAIEDANN:SRANAGSLSEPMARASTREX
EAAS:YERYOARLTTFNVDFDCLRLPVLQLEAMEDIWGWRRIRGYLVDECOOTN
CAGY:CLKMLAGPRGNFTVCDDOS:VAMGAPENLOQVARDYPAJEI:KLEQNYR
CSNRV:RAANALIAHNSHEH:KTLMSDADGERIRVWECRDEHEAKVAEISLGT
KQCVKSPCILFCNFSRPLEKALQ:LRVPHYLTGTATLERGEVDLLSWRLIV
NPDQAFIRVQSPREVGAT:ARLAEASAKSVPMRAESMGALOHUPPRAANG
LSAFTLILRXHREKSSVPAGELVRLADKSLGLNDRNSKDETGFQRKXNDELA
EMFEGPRGASQALALLSRNCKDDGQVMMTHASKGLEFRYFVIVGCEG
VLPHEVLEGNLOERELLVYGT:TRAKEQ:WMSHSL:TRKPGEHVRLKPRFFDEIP
AEELQDGDGPFVADAEKKEKASAGLAIQALFD"
5341..5970
/gene="tdk"
/note="synonym: XAC4282"
5341..5970
/gene="tdk"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="tyrosine kinase"

```

## TITLE

## JOURNAL

## Submitted (28-NOV-2001): Departamento de Bioquímica, Universidade de

## Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,

## Brazil

## Location/Qualifiers

## 1..11004

## /organism="Xanthomonas axonopodis pv. citri str. 306"

## /mol\_type="genomic DNA"

## /strain="306"

## /db\_xref="taxon:90486"

## /note="pathovar: Citri"

## 211..679

## /gene="XAC4278"

## 211..879

## /gene="XAC4279"

## /note="identified by sequence similarity; putative; ORF

## located using Blastx/Glimmer/Genemark"

## /codon\_start=1

## /transl\_table=1

## /product="conserved hypothetical protein"

## FEATURES

## source

```

/protein_id="AA039117.1"
/db_xref="GI:21110723"
/translation="MAKLVYSSAMNAGKTTTLCSAHNYRVRGMSLTLITPKLDHRA
GSGVVASRIGLRADCRIFEDTQLCOQVEDIHNDGALHCYLVDSERAFQFUSRAQVWYLS
EVVDRKIFVLICVGRVDFRFEGLFEGSCFLWAGDELEIKTCHGSKATYTVRVDL
HGHAQVQGPVVEIGNERYVSRAEFKKIRNGEGRIDPLQIALPLVA"
gene
6391..7647
CDS
6391..7647
/gene="XAC4283"
6391..7647
/gene="XAC4283"
/notes="identified by sequence similarity; putative; ORF
located using BlastX/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AA039118.1"
/db_xref="GI:21110724"
/translation="MEALXDPPLPALRADARVGRSSVDPVLSVLTITLITGMSFAVVA
RVTDARWITQVHDAISFGFDEPLATTFDCSVVRGNVAFHGASDPTFYEDHP
SPRYGESVSVPIKEDDGSVGTICALDQPRVMDAPLVEXVELLAGLLAAQIOAE
CRAPESALSRARSEKRGVAGARLSEEFIGLHDLRPLQSMHVAVMLSIQOLN
HRYGSSRLSQRSLRVEELIDPACDPARGIEIDMLQLDYDGSDDLDAISOVDETR
AAYPNQVLSHRAIDEPVHGDARLAQNGFSLKNVAVQPSQLKAVAVTEHGRLR
TEVNDGIDPRVLDVHASTHVTHTGGRALPDVDLGRVAAQIAKARQSGEIRIISKO
GTGFRVSLACTRPRGRHLVGTGAPT"
complement(7814..9415)
/gene="mdog"
/notes="synonym: XAC4284"
complement(7814..9415)
/gene="mdog"
/notes="identified by sequence similarity; putative; ORF
located using BlastX/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="periplasmic glucan biosynthesis protein"
/protein_id="AA039119.1"
/db_xref="GI:21110725"
/translation="MQRHFELKNAALAAALGPTLPQWALAAKAYGVJRLQGPQPPD
YMLKGGARLAPKPYKSHQLPLGFEALNMDQVGSIRYRQDHALWGNKQKFOAK
FHLGLYFHTVPIVEIVDKAQQLAVDPAFVDSRGLGKQJFKDLGAGPLNTRK
CTDPEFSAFGASVFAVRNGVSGSARGLAIDTGTGSEEPEDIAVYLQEFALFS
DTVYVGLDPSVACVARRAINGSVLWLDISALYPRKALRGLIGRTSYQGE
NKNKMDRPEHDTDGLAMWCGGEMWRPJCNPFLRNFVDENGRGLLQRC
RNDHYQGVFVEKPLCPKRWKSGVQVLEIPTVCTEDVIAFWNPQKPC
FGSLLXGLYLGAQCPASSPAHCVATRTGGIVGQKRSFHSFAYVQAGSALA
ALAKDPKAYEALVQSRGTEVSAREPLHELKGVAMFDLVDPDSGTCQIDIRLYR
ANGKPLTETWYQWTPPPPSRRY"
complement(9494..10306)
/gene="XAC4285"
complement(9494..10306)
/gene="XAC4285"
/notes="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA039120.1"
/db_xref="GI:21110726"
/translation="MRPNRSGSLLETVETSPNALRPGQLQADIARTGTWPAATCOQA
7.2%; Score 19; DB 1; Length 11004;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Query Match
216 GTTCGGAGCCTGATGATC 234
|||||
Db 7332 GTTCGGAGCCTGATGATC 7320

```

```

RESULT 9
LOCUS
DEFINITION
Xanthomonas campestris pv. campestris str. ATCC 33913, section 448
of 460 of the complete genome.

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PubMed
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS

```

AE012540 AE009322  
AE012540.1 GI:21115424

Xanthomonas campestris pv. campestris str. ATCC 33913  
Xanthomonas campestris pv. campestris str. ATCC 33913  
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
Xanthomonadaceae; Xanthomonas.

1 (bases 1 to 11854)  
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Parah, C.S., Furlan, L.R.,  
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
Camargo, L.E.A., Camarotte, G., Canavan, F., Cardoso, J.,  
Chambergo, F., Clapina, I.P., Cicatelli, R.M.B., Coutinho, L.L.,  
Cursino-Santos, J.R., E-Dorrry, H., Ferreira, J.B., Ferreira, A.J.S.,  
Ferreira, R.C.C., Ferro, V.L., Formighieri, E.F., Franco, M.C.,  
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
Madelra, A.V.B.N., Martinez-Rossi, N.Y., Martins, E.C., Meidanis, C.,  
Venck, C.F.X., Miyaki, C.Y., Koch, D.H., Moreira, L.M., Novo, M.T.M.,  
Okura, V.K., Oliveira, N.C., Oliveira, V.R., Pereira, C.H.A.,  
Rosi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, J.A.F.,  
Takita, M.A., Tanura, R.E., Teixeira, E.C., Tezza, R.O., Trindade dos  
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
Kitajima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with  
differing host specificities  
Nature 417 (6887), 459-463 (2002)

2222145  
1284217

2 (bases 1 to 11854)  
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Parah, C.S., Furlan, L.R.,  
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
Camargo, L.E.A., Camarotte, G., Canavan, F., Cardoso, J.,  
Chambergo, F., Clapina, I.P., Cicatelli, R.M.B., Coutinho, L.L.,  
Cursino-Santos, J.R., E-Dorrry, H., Ferreira, J.B., Ferreira, A.J.S.,  
Ferreira, R.C.C., Ferro, V.L., Formighieri, E.F., Franco, M.C.,  
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
Madelra, A.V.B.N., Martinez-Rossi, N.Y., Martins, E.C., Meidanis, C.,  
Venck, C.F.X., Miyaki, C.Y., Koch, D.H., Moreira, L.M., Novo, M.T.M.,  
Okura, V.K., Oliveira, N.C., Oliveira, V.R., Pereira, C.H.A.,  
Rosi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, J.A.F.,  
Takita, M.A., Tanura, R.E., Teixeira, E.C., Tezza, R.O., Trindade dos  
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
Kitajima, J.P.

Direct Submission  
Submitted (23-NOV-2001) Departamento de Bioquímica, Universidade de  
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508 900,  
Brazil

Location/Qualifiers  
1..11854  
/organism="Xanthomonas campestris pv. campestris str. ATCC  
33913"  
/molecule\_type="genomic DNA"  
/strain="ATCC 33913"  
/db\_xref="ATCC:33913"  
/db\_xref="taxon:190485"  
/notes="pathovar: campestris"  
199..1446  
/gene="XCC4144"  
199..1446  
/gene="XCC4144"  
/notes="identified by sequence similarity; putative; ORF  
located using BlastX/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="sensor histidine kinase"  
/protein\_id="AA043365.1"  
/db\_xref="GI:21115427"

translation="MGOLPSELHADVARVGLSSVPVLTLLTETGCRANAVARVTD  
ARWTCVQVQLAFSLQPGRELPLATTFDCGVRNRQAVSFGHAFVYDPRFSPKL  
YGFSESVFVRUDSGVFGTQLQDPEPRVMDAPLVEXVELLAGLLAAQIALRAB

```

OSALESQARSELGRVCASARLREEFIGILCHLRLXPLOSMHVVDMLS'DOLSRQ3
NQVLSQRSRRMEEDCFASDFARG-ERQWLQDYGNGSGLLDLALSQVDETRAAYP
NQVLSARVADEPVHDAIRIAQMFGLMINAVVQOOSLIIKAIATVEGGRIRIEVC
NLGIEPRRCDILHASTVHTAGRALPEVDGLFMAAQIAQAHQGBELRTSGKDGTCF
HVELACORPRARMQVDRNSTPA"
complement (1474..3075)
/gene="cdog"
/note="synonym": XCC4145"
complement (1474..3075)
/gene="ndog"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1;
/trans_table=1;
/product="periplasmic glucan biosynthesis protein"
/protein_id="AA043366.1"
/db_xref="GI:21115428"
/translation="MQRHPTQVAAALALGLPALPPWALAARAVGLRPLGQFPQPF
YAMKGOARALAKAFYKSHQVLPGLPESNTQYQSIYRQDHALWADNGKEQAKP
FHLGLYHTPHVYEDVGRQQLAYDPAAFYGRSLGKQKPLKDLFAGFRLNTRK
DIDFSAFLGASTFRAVYKGGQGGARGLAIDTG'GGPEEPDFDIATYLEQPADDS
DTVVYGLJDSPSVSGAYRAITNGEVLVMDISALYPRKAIERLIGICTSQTGCE
NDRKMDWRPEIHTDGLAWTGGGEWIWRPLCNPHLRFNMFVDENPRGFGLLGRD
RNFHYQDQGVYKREKPLWEPKSGMGKSVOLVETPTVDETFDNIIVAFWNPQAKPO
PGOELLKRYLVGLWAPASSPLAHCVATRTGLGGVGGKRSFHSFRAVDPAAGELA
ALAKDPKAYEVLQVSRGTETIVSARPLHELKYRAMFPLVPPDEGTQCIDIRLPER
ANGKPLTETWLYQWTFPPASERKIY"
complement (3148..4266)
/gene="XCC4146"
complement (3148..4266)
/gene="XCC4146"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1;
/trans_table=1;
/product="conserved hypothetical protein"
/protein_id="AA043367.1"
/db_xref="GI:21115429"
/translation="MDALEQVAKVMPFVAVVTGLVLPSSAGCACTARKTERALZ
RLILGNVIEVATPSPFGGILAFQLRQSPVLKLGNAVEQQAPQVSGTAGDIAVLG
HDVWVGQVSHWJQQVQPKAKAGAVPPPEYLSATTATITEDATQVMGVPSPV
TGVLKRSALVSREPTIDVQATARNRTPELSWDLWENTRVAAATRVFVVDAAAD
LRQPPDQGTGTPVLEGLSLFVLSADTADTLIQKELLOPSAGMAGFAGTQVL
VIRPAHQPKARIRHPCQVELYLADPPCHPEHGLEWEHAPYRTLAPGDMQAOQEQW
TLLEYTGENROAQRQFLCSKACELATNACAGSPAVP"
complement (4953..5318)
/gene="mutM"
/note="synonym": XCC4147"
complement (4953..5318)
/gene="mutM"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1;
/trans_table=1;
/product="formamido pyrimidine DNA glycosylase"
/protein_id="AA043368.1"
/db_xref="GI:21115430"
/translation="MPELPEVETTLRGLAPLVCORHTGVILRRPOLRWP'RAQIECL
LEGATITQVRPAKYLDITDAGSAYHLGMSGLVPLRGDTPPRAHWHVDSJNG
RVLFNDRPRGCLLWQDRETHELLASLPFELSAFTGYJHJACAGRAAVKIFL
MQQAVVGVGNLYAASCHRAQSPLEAGKVRERYRRLADAVKELAYALRGSTT
LRDIFSDGAPGYGEQELMYGREGACRHCGBELKHATIGQRATWYCAACQR"
complement (6263..6381)
/gene="XCC4148"
complement (6263..6381)
/gene="XCC4148"
/product="SSU 5S ribosomal RNA"
/note="operon 2"
complement (6507..9389)
/gene="XCC4149"
complement (6507..9389)
/gene="XCC4149"
/product="SSU 23S ribosomal RNA"
/note="operon 2"

```

```

complement (9617..9693)
/gene="XCC4150"
complement (9617..9693)
/gene="XCC4150"
/product="RNA-11e"
/note="Found by RNAscan"
complement (9713..9789)
/gene="XCC4151"
complement (9713..9789)
/gene="XCC4151"
/product="rRNA-A1a"
/note="Found by RNAscan"
complement (9883..11429)
/gene="XCC4152"
complement (9883..11429)
/gene="XCC4152"
/product="SSU 16S ribosomal RNA"
/note="operon 2"
BASE COUNT 2360 a 3765 c 3189 g 2540 t
ORIGIN
Query Match. 7.2%; Score 13; DB 1; Length 11854;
Best Local Similarity 100.0%; Pred.No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 216 GTTCGGCAGCTGATGATC 234
|||||
Db 1098 GTTCGGCAGCTGATGATC 1116
|||||

RESULT 10
HS3646P11
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT
HS3646P11 14616 bp DNA linear PRI: 21-JAN-2003
Human DNA sequence from clone RP4-646p11 on chromosome 1, comp.ete
sequence.
AL049715
AL049715.26 GI:2787819
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Wallis.
1. (bases 1 to 14616)
Direct Submission
Submitted (21-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB11 8SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: c.onerequest@sanger.ac.uk
On Jan 21, 2003 this sequence version replaced gi:9581791.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Embl, ENMBL, Sw, Swissprot, Tr, TREMBL, Wp, WORMPEP, Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

```





was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>  
 RP11-114018 is from the library RP11-11.1 constructed by the group of Peter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-114018. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RP11-114018 is at 1070356 in this sequence. The true right end of clone RP11-323X8 is at 2002 in this sequence.

#### FEATURES

source

Location/Qualifiers  
 1..1070356  
 /organism="Homo sapiens"  
 /ol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-114018"  
 /clone\_lib="RP11-11.1"

BASE COUNT 32752 a 22325 c 26837 g 31138 t

#### Query Match

7.2% Score 19 DB 9; Length 1070356;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 20 AGCCAGCAGCCAGCCAGTT 38

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

#### REFERENCE

AUTHORS

TITLE

JOURNAL

Worley, K.C.

Direct Submission

Submitted (29-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

Sequence format modified by NCBI staff

On Jan 29, 2003 this sequence version replaced gi:6056152.

NOTE: This is a 'working draft' sequence. It currently

consists of 52 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

820: contig of 820 bp in length

821: contig of 120 bp

822: contig of 1072 bp in length

1391: contig of 120 bp

2393: contig of 1246 bp in length

3339: contig of 100 bp

3439: contig of 1123 bp in length

4562: contig of 100 bp

4662: contig of 838 bp in length

5519: contig of 100 bp

5520: contig of 1450 bp in length

5620: contig of 100 bp

7070: contig of 832 bp in length

8102: contig of 100 bp

8327: contig of 826 bp in length

8328: contig of 1034 bp in length

10362: contig of 100 bp

10362: contig of 14 bp in length

10375: contig of 120 bp

10976: contig of 581 bp in length

11557: contig of 100 bp

11557: contig of 935 bp in length

11557: contig of 100 bp

12492: contig of 100 bp

12492: contig of 846 bp in length

13338: contig of 100 bp

13338: contig of 1834 bp in length

13372: contig of 120 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

13372: contig of 100 bp

```

* 37749 39440: contig of 1692 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2220 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 1750 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2707 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2215 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2596 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2459 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2534 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2756 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 3135 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2070 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 4177 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 2996 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 3029 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 5107 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 5027 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 5922 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 6322 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 7537 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 7412 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 7412 bp in length
* 37749 39440: gap of 100 bp
* 37749 39440: contig of 11794 bp in length.
* 37749 39440: Location/Qualifiers
* 37749 39440:
* 37749 39440:   organism="Drosophila melanogaster"
* 37749 39440:   mol_type="genomic DNA"
* 37749 39440:   db_xref="taxon:7227"
* 37749 39440:   /chromosome="3L/79A3"
* 37749 39440:   /clone="RPc198-2503"
* 37749 39440: 35760 a 29458 c 29103 g 36584 t 5150 others
* 37749 39440:
* 37749 39440: Query Match 7.2%; Score 19; DB 2; Length 135055;
* 37749 39440: Best Local Similarity 100.0%; Pred. No. 1.2e+02;
* 37749 39440: Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
* 37749 39440:
* 37749 39440: QY 78 CAGCTGACGACGACGCTG 96
* 37749 39440:
* 37749 39440: DE 82577 CAGCTGACGACGACGCTG 82559
* 37749 39440:
* 37749 39440: RESULT 15
* 37749 39440: LOCUS AC011909/c 135055 bp CNA linear HTG 29-JAN-2003
* 37749 39440: DEFINITION Drosophila melanogaster chromosome 3L/79A3 clone RPc198-2503. ***
* 37749 39440: SEQUENCING IN PROGRESS ***; 52 unordered pieces.
* 37749 39440: AC011909
* 37749 39440: AC011909.2 GI:28091257

```

```

HTG, HTGS, PHASE:
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 bases 1 to 135055:
Muzny D.M., Adams C., Bailey K., Barbara J., Blankenburg K.,
Bodora B., Bouck C., Bowie S., Brooks A., Euhay C., Euhay C.,
Burkett C., Burrows J., Carter M., Chacko J., Chen Z., Cox C.,
David R., Delgado C., Deshazo D., Ding Y., Domah-Rashid N.,
Dugan-Rocha S., Dubin K.J., Fernandez C., Ferraguto D.,
Focur-Tarsey J., Frantz P., Gresh R., Gorrell J., Gorrell L.J.,
Guerara K., Harris K., Hernandez J., Hodgson A., Hughes M.,
Hollway C., Kozak M., Jackson D., Jackson D., Cia Y., Jones M.,
Kelly S., Kondejewski N., Kong Y., Kovar C., Lea P., Li Z.,
Lichtarge C., Liu J., Liu W., Lohman G., Liu C., Lucier F.,
Martin R., Martinez C., McLeod X.P., Mei G., Morgan M., Morris S.,
Nash S., Nelson A., Nguyen R., Nguyen L., Perez L., Pt J., Quiles M.,
Pattish B., Paxton S., Payton B., Perez L., Pt J., Quiles M.,
Reiter D., Rives M., Samuel S., Say J., Scherer S., Shah E.,
Shen H., Simon M., Sparks A., Stamps A., Sugang R., Taber P.,
Taylor T., Vasquez L., Vinson R., Vo Q., Warban M., Washington S.,
Wrenstock G., Weinstock G., Weinstock I., Williams A., Worley K., Wren J.,
Wrenstock G., Yu W., Zhou X., Nelson D. and Gibbs R.
Direct Submission
Unpublished
2 bases 1 to 135055:
Worley K.C.
Direct Submission
Submitted (16-Oct-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
1 bases 1 to 135055:
Worley K.C.
Direct Submission
Submitted (29-JAN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
Sequence format modified by NCI staff
On Jan 29, 2003 this sequence version replaced gi:28091257.
* NOTE: This is a working draft sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 820: contig of 920 bp in length
* 920: gap of 100 bp
* 921: contig of 1372 bp in length
* 1993: gap of 100 bp
* 2093: contig of 1246 bp in length
* 3139: gap of 100 bp
* 3439: contig of 123 bp in length
* 4561: gap of 100 bp
* 4562: gap of 100 bp
* 4562: gap of 100 bp
* 5519: contig of 838 bp in length
* 5520: gap of 100 bp
* 5620: contig of 1450 bp in length
* 7070: gap of 100 bp
* 7070: contig of 832 bp in length
* 8002: gap of 100 bp
* 8002: contig of 826 bp in length
* 8028: gap of 100 bp
* 9028: contig of 1034 bp in length
* 10062: gap of 100 bp
* 10162: contig of 714 bp in length
* 10876: gap of 100 bp
* 10976: contig of 581 bp in length
* 11657: gap of 100 bp
* 11657: contig of 835 bp in length
* 12492: gap of 100 bp

```

```
* 12592 13437: contig of 846 bp in length
* 13438 13537: gap of 100 bp
* 12538 13771: contig of 1834 bp in length
* 15332 15471: gap of 100 bp
* 15472 17147: contig of 1676 bp in length
* 17148 17247: gap of 100 bp
* 17248 18753: contig of 1506 bp in length
* 18754 18853: gap of 100 bp
* 18854 18978: contig of 1025 bp in length
* 18979 21419: contig of 1441 bp in length
* 21420 21519: gap of 100 bp
* 21520 22834: contig of 1115 bp in length
* 22835 22934: gap of 100 bp
* 22935 24335: contig of 1501 bp in length
* 24336 24535: gap of 100 bp
* 24536 26082: contig of 1547 bp in length
* 26083 26182: gap of 100 bp
* 26183 27416: contig of 1234 bp in length
* 27417 27516: gap of 100 bp
* 27517 29266: contig of 1750 bp in length
* 29267 29366: gap of 100 bp
* 29367 31249: contig of 1982 bp in length
* 31249 31348: gap of 100 bp
* 31349 32698: contig of 1350 bp in length
* 32699 32798: gap of 100 bp
* 32799 33863: contig of 1065 bp in length
* 33864 33963: gap of 100 bp
* 33964 36244: contig of 2280 bp in length
* 36244 36343: gap of 100 bp
* 36344 37648: contig of 1305 bp in length
* 37649 37748: gap of 100 bp
* 37749 39440: contig of 1692 bp in length
* 39441 39540: gap of 100 bp
* 39541 41760: contig of 2220 bp in length
* 41761 41860: gap of 100 bp
* 41861 43610: contig of 1750 bp in length
* 43611 43710: gap of 100 bp
* 43711 46417: contig of 2707 bp in length
* 46418 46517: gap of 100 bp
* 46518 48732: contig of 2215 bp in length
* 48733 48832: gap of 100 bp
* 48833 51428: contig of 2596 bp in length
* 51429 51528: gap of 100 bp
* 51529 51987: contig of 2459 bp in length
* 51988 54087: gap of 100 bp
* 54088 56621: contig of 2534 bp in length
* 56622 56721: gap of 100 bp
* 56722 59477: contig of 2756 bp in length
* 59478 59577: gap of 100 bp
* 59578 62712: contig of 3135 bp in length
* 62713 62812: gap of 100 bp
* 62813 64882: contig of 2070 bp in length
* 64883 64982: gap of 100 bp
* 64983 69159: contig of 4177 bp in length
* 69160 69259: gap of 100 bp
* 69260 72255: contig of 2396 bp in length
* 72256 72355: gap of 100 bp
* 72356 75159: contig of 3004 bp in length
* 75160 75259: gap of 100 bp
* 75260 78631: contig of 3172 bp in length
* 78632 78731: gap of 100 bp
* 78732 81527: contig of 2796 bp in length
* 81528 81627: gap of 100 bp
* 81628 84656: contig of 3029 bp in length
* 84657 84756: gap of 100 bp
* 84757 89863: contig of 5107 bp in length
* 89864 89963: gap of 100 bp
* 89964 94990: contig of 5027 bp in length
* 94991 95090: gap of 100 bp
* 95091 100921: contig of 5902 bp in length
* 100922 101021: gap of 100 bp
* 101022 107412: contig of 6320 bp in length
```

```
* 107413 107512: gap of 100 bp
* 107513 115449: contig of 7937 bp in length
* 115450 115459: gap of 100 bp
* 115500 123161: contig of 7612 bp in length
* 123162 123261: gap of 100 bp
* 123262 135055: contig of 11794 bp in length.
```

#### FEATURES

```
Location/Qualifiers
    ..135055
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
    /db_xref="taxon:7227"
    /chromosome="3L:79A3"
    /clone="RP1193-253"
```

BASE COUNT 35760 a 28456 c 29103 g 36584 t 1150 others  
ORIGIN

Query Match 7.2% Score 19, DB 2: Length 135055;  
Best local similarity 100.0%, Pred. No. 1: 2e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 CAGCCTGAACGAGCAGCTG 96

Db 82577 CAGCCTGAACGAGCAGCTG 82559

Search completed: November 7, 2003, 14:25:12  
Job time: 1517 secs